École des Ponts ParisTech

PhD thesis

École des Ponts ParisTech

Registration of Heterogenous Data for Urban Modeling

Doctoral School: Mathématiques et Sciences et Technologies de l'Information et de la Communication

Domain: Computer Science

Thesis defended on June, 2022, by
Rahima DJAHEL

Jury composition:

Beatriz Marcotegui Prof, Mines Paris Tech	*Reviewer*
Tania Landes HDR, INSA Strasbourg	*Reviewer*
Paul Checchin Prof, Université Blaise Pascal	*Examiner*
Nicole Vincent Prof, Université Paris Descartes	*Examiner*
Raouf Ben Jemaa Engineering Director, Trimble	*Examiner*
Pascal Monasse Prof, École des Ponts ParisTech	*Supervisor*
Bruno Vallet HDR, Institut National de l'Information Géographique et Forestière	*Supervisor*

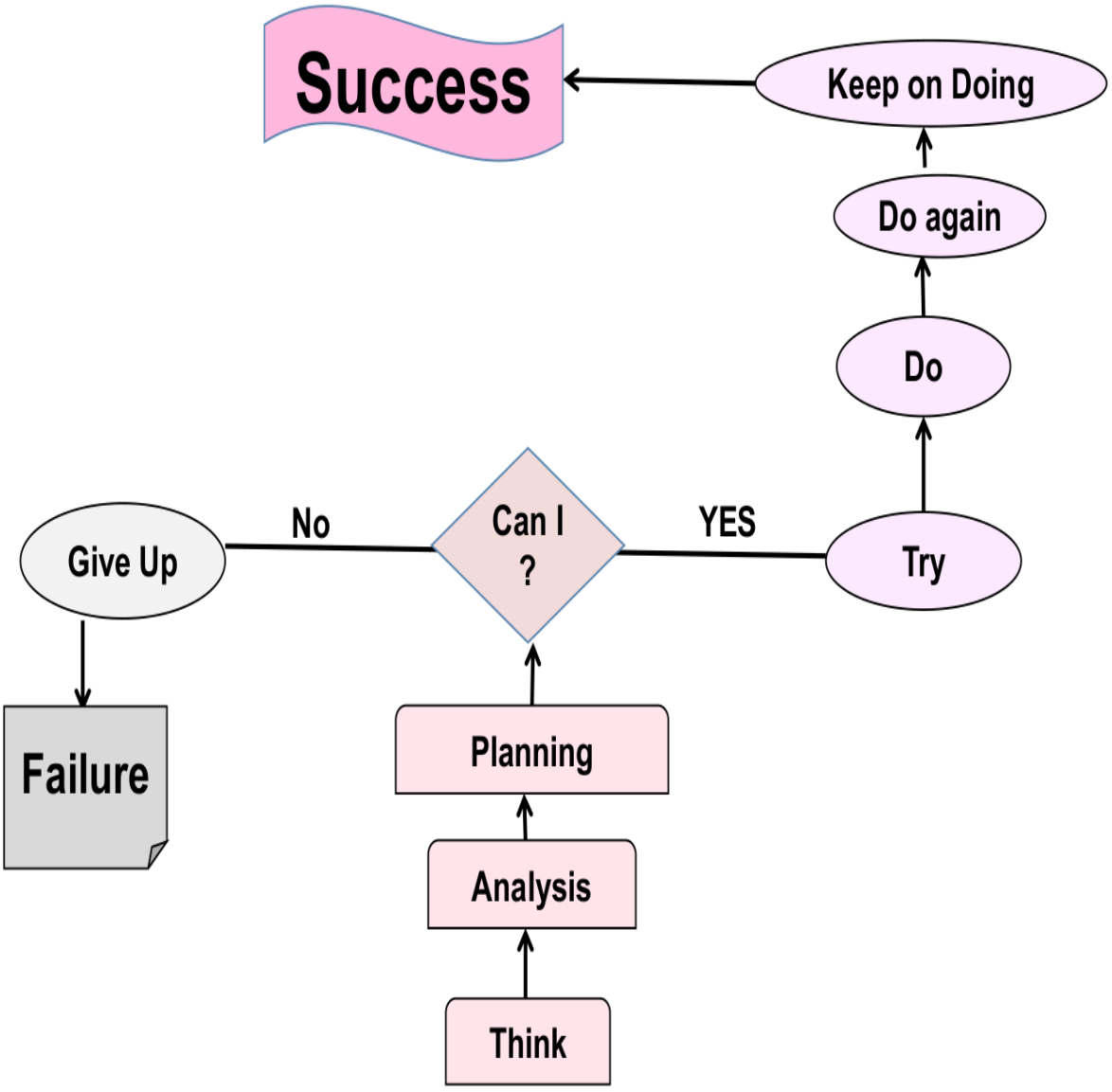


Figure 1: Success Diagram

Contents

List of Tables

List of Figures

Acronyms

ALS Airborne Laser Scanning.

BIM Building Information Modeling.

BIOM Building Indoor/Outdoor Modeling.

CCD Charge Coupled Device.

CMOS Complementary Metal Oxide Semiconductor.

HT Hough Transform.

ICL Iterative Closest Line.

ICP Iterative Closest Point.

LiDAR Light Detection and Ranging.

LSD Line Segment Detector.

MI Mutual Information.

MLS Mobile Laser Scanning.

MLSD multiscale line segment detector.

MMS Mobile Mapping System.

MSAC M-estimator Sample Consensus.

NDT Normal Distribution Transform.

NFA Number of False Alarms.

PCA Principal Component Analysis.

RANSAC Random Sample Consensus.

RG Region Growing.

SA Simulated Annealing.

SfM Structure From Motion.

TLS Terrestrial Laser Scanning.

UAV Unmanned Aerial Vehicle.

Abstract

Indoor/Outdoor modeling of buildings is an important issue in the field of building life cycle management. It is seen as a joint process where the two aspects collaborate to take advantage of their semantic and geometric complementary. This global approach will allow a more complete, correct, precise and coherent reconstruction of the buildings. This thesis is part of the Building Indoor/Outdoor Modeling (BIOM) ANR project that aims at automatic, simultaneous indoor and outdoor modelling of buildings from image and dense point clouds. The first ambition of the BIOM ANR project is to integrate heterogeneous data sources for buildings modeling. The heterogeneity is both in: data type (image/ LiDAR data), acquisition platform (Terrestrial/ Aerial), acquisition mode (dynamic/static) and point of view (indoor/outdoor). The first issue of such modeling is thus to precisely register this data. The work carried out has confirmed that the environment and the type of data drive the choice of the registration algorithm. Our contribution consists in exploiting the physical and geometric properties of the data and the acquisition platforms in order to propose potential solutions for all the registration problems encountered by the project. As in a building environment, most objects are composed of geometric primitives (planar polygons, straight lines, openings), we chose to introduce registration algorithms based on these primitives. The basic idea of these algorithms consists in the definition of a global energy between the extracted primitives from the data-sets to register and the proposal of a robust method for optimizing this energy based on the RANSAC paradigm. Our contribution ranging from the proposal of robust methods to extract the selected primitives to the integration of these primitives in an efficient registration framework. Our solutions have exceeded the limitations of existing algorithms and have proven their effectiveness in solving the challenging problems encountered

by the project such as the indoor (static mode)/outdoor (dynamic mode) registration, image/LiDAR data registration, and aerial/terrestrial registration.

Key words

Registration, Geometric primitives, Planar polygons, 3D Segments, Openings, Global robust energy, Optimization, RANSAC.

Résumé

La modélisation intérieure/extérieure des bâtiments est un enjeu important dans le domaine de la gestion du cycle de vie des bâtiment. Elle est vue comme un processus conjoint où les deux aspect collaborent pour tirer partie de leurs complémentarités sémantiques et géométriques. Cette approche globale permettra une reconstruction complète, correcte, précise et cohérente des bâtiments. Cette thèse s'inscrit dans le cadre du projet ANR BIOM (Modelisation Intérieur/Extérieur de Bâtiments) qui vise à la modélisation automatique et simultanée de l'intérieur et de l'extérieur de bâtiments à partir d'images et de nuages de points denses. La première ambition du projet ANR BIOM est d'intégrer des sources de données hétérogènes pour la modélisation des bâtiments. L'hétérogénéité est à la fois dans :

- Le type de données (image/données LiDAR).
- La plate-forme d'acquisition (Terrestre/Aérienne).
- Le point de vue (intérieur/extérieur).
- Le mode d'acquisition (dynamique/statique)

Les approches de l'état de l'art traitent en général soit l'extérieur, soit l'intérieur et se limitent alors souvent à une seule pièce, au mieux un seul étage, utilisant des a priori forts de parallélisme et d'orthogonalité qui ne sont pas nécessairement vérifiés. De notre point de vue, aucune méthode n'a proposé une modélisation conjointe intérieur/extérieur dans le cadre d'un formalisme unifié. La numérisation d'un objet à partir d'une seule station d'acquisition est souvent difficile. Pour couvrir toutes les faces de l'objet, plusieurs points de vue différents sont nécessaires. Cela est dû à des raisons diverses, soit le champ de vision du scanner est limité, soit l'objet est de dimension ou d'architecture complexe, qui pourrait être aussi un

obstacle empêchant une visée directe et génèrant des masquages sur la surface à scanner. Les nuages de points provenant de ces stations multiples sont définis chacun dans un repère relatif au scanner laser. Un regroupement dans un même système de coordonnées de l'ensemble des nuages de points acquis depuis les différentes stations est nécessaire. Ce processus s'appelle recalage ou encore consolidation. Les travaux menés ont confirmé que l'environnement et le type de données conditionnent le choix de l'algorithme de recalage. Notre contribution consiste à exploiter les propriétés géométriques et physiques des données et des plateformes d'acquisition afin de proposer des solutions potentielles à tous les problèmes de recalage rencontrés par le projet. Comme dans un environnement de bâtiments la plupart des objets sont composés de primitives géométriques (polygones planaires, lignes droites, ouvertures), nous avons choisi d'introduire des algorithmes de recalage basés sur ces primitives. L'idée de base de ces algorithmes consiste en la définition d'une énergie globale entre les primitives extraites à partir des jeux de données à recaler et la proposition d'une méthode robuste pour optimiser cette énergie basée sur le paradigme RANSAC. Nos contributions vont de la proposition de méthodes robustes pour l'extraction des primitives sélectionnées à l'intégration de ces primitives dans un cadre de recalage efficace. Le recalage intérieur/extérieur avec une précision proche de celle d'un scan (environ 1 cm) représente un défi majeur pour la modélisation intérieure/extérieure des bâtiments. L'absence de chevauchement entre les données intérieures et extérieures est l'obstacle le plus souvent rencontré, d'autant plus lorsque les deux ensembles de données sont acquis séparément et à l'aide de différents types de capteurs. Pour faire face à ce problème, nous avons proposé deux solutions initiales puis nous les avons combinées. La première solution repose sur la détection et la mise en correspondance des polygones planaires. Le point fort de cette solution réside dans le fait qu'elle exploite le très faible chevauchement entre les scans intérieurs et extérieurs du même bâtiment en détectant les points situés à l'intérieur des bâtiments lors des scans extérieurs comme des points où le rayon laser traverse les façades détectées à travers les ouvertures. Cette solution définit une énergie globale robuste entre deux ensembles de polygones et propose une méthode robuste pour maximiser cette énergie basée sur le paradigm RANSAC. La maximisation de cette énergie conduira à une maximisation de chevauchement et

minimisation de la distance sur ces chevauchements. Afin de simplifier cette maximisation, nous avons procédé à une étape de regroupement des polygones planaires de chaque ensemble selon la direction de la normale. Ce regroupement a généré trois clusters directionnels (un horizontal et deux verticaux). L'association de ces clusters nous a permis de trouver la rotation optimale. La génération et l'évaluation des hypothèses (des correspondances possibles de polygones) nous ont permis de trouver la translation optimale. La deuxième solution initiale repose sur la détection et la mise en correspondance des ouvertures. Comme une ouverture est considérée parmi les entités communes qui peuvent être vues de l'intérieur et de l'extérieur, elle peut aider au recalage des nuages de points intérieurs et extérieurs. Le recalage est généralement effectué en faisant correspondre des caractéristiques ou bien en optimisant une energie globale entre les données à recaler. Dans notre cas, les caractéristiques sont deux ensembles d'ouvertures verticales détectées à partir de scans intérieurs et extérieurs. Cependant, ces ouvertures ne sont pas assez caractéristiques pour permettre les mises en correspondance de manière robuste. Comme une ouverture est définie par une forme rectangulaire composée de quatre segments, deux horizontaux et deux verticaux, nous avons modélisé notre problème de recalage comme une minimisation d'une energie globale robuste entre deux ensembles de segments et nous avons proposé une méthode robuste pour minimiser cette energie suivant le paradigme RANSAC. Sachant que la minimisation de cette énergie va entraîner une maximisation de chevauchement et minimisation de la distance sur ces chevauchements.

Les deux solutions initiales sont très efficaces pour réaliser le recalage intérieur/extérieur, mais elles ont quelques limitations. La solution basée sur des polygones planaires a une incertitude dans la direction horizontale parallèle à la façade, tandis que la solution basée ouvertures a une incertitude dans la direction orthogonale à la façade. Donc, afin de supprimer ces incertitudes, nous avons proposé une solution basée sur des associations de polygones planaires et d'ouvertures. Cette nouvelle solution a prouvé sa performance pour recaler les scans intérieurs et extérieurs. L'analyse et la reconstruction de scènes 3D à partir d'images et données LiDAR est un domaine de recherche actif en vision par ordinateur. D'une part, les données Li-

DAR fournissent des informations de surface très précises et robustes. D'autre part, l'image fournit des détails à haute résolution, mais la géométrie 3D estimée à partir de l'appariement dense est moins robuste et précise. Par conséquent, l'intégration des données de ces deux sources peut conduire à une segmentation sémantique robuste et une reconstruction complète de scènes 3D. Donc, un autre objectif de cette thèse consiste à résoudre le problème du recalage image/données LiDAR.

Ce problème est très difficile compte tenu de la très forte différence de modalité. Cette différence rend difficile l'extraction de caractéristiques comparables entre les deux modalités. À l'inverse, avec nos primitives et notre approche basée sur RANSAC, nous supposons que la meilleure transformation est celle qui optimise une energie globale robuste entre toutes les primitives extraites et donc qui est la plus cohérente avec toutes les informations des jeux de données à recaler. Notre algorithme de recalage basé sur des segments 3D représente une solution potentielle à ce genre de problème. Cet algorithme est basé sur la minimisation d'une énergie globale robuste, qui a été définie dans l'algorithme de recalage basé ouvertures, entre deux ensembles de segments 3D. Cette minimisation est basée sur le paradigme RANSAC. Afin de simplifier cette minimisation, nous avons commencé par regrouper les segments 3D de chaque jeu de données selon leurs direction. Les clusters obtenus sont associés pour trouver les rotations possibles, puis les segments 3D des clusters associés sont mis en correspondance afin de trouver le facteur d'échelle et la translation minimisant l'énergie définie. Nous avons choisi de travailler avec des segments 3D pour résoudre ce problème parce que dans un environnement bâti, il semble que la meilleure caractéristique à utiliser pour le recalage image/données LiDAR soit le segment 3D. Les segments 3D peuvent être extraits de manière fiable, précise et automatique à la fois à partir des données LiDAR et à partir des données image surtout si nous avons un chevauchement suffisant d'images. Elles agrègent plus d'informations que les points, sont donc moins sensibles au bruit et sont plus fréquentes que les primitives plus complexes (rectangles). L'objectif de cette thèse est non seulement le recalage des données hétérogènes (image et données LiDAR) mais aussi le recalage des données acquises par des plateformes d'acquisition hétérogènes (aériennes/terrestres). En utilisant des plateformes aéroportées au milieu urbain,

on peut voir les faces supérieures des bâtiments, mais les faces latérales manquent toujours. Contrairement aux plateformes terrestres qui donnent des informations complètes et denses sur les faces latérales et ne donnent pas d'informations significatives sur les faces supérieures. Donc, afin d'obtenir une couverture complète des zones urbaines, la fusion des vues aériennes et terrestres est nécessaire. Les données aériennes peuvent être acquises sous forme d'images ou des données LiDAR et c'est le même cas pour les données terrestres. Donc le problème du recalage aérien/terrestre peut prendre plusieurs formes :

- Recalage du LiDAR aérien et LiDAR terrestre en utilisant soit l'algorithme basé sur les polygones planaires soit l'algorithme basée sur des segments 3D.
- Recalage du LiDAR aérien et l'image terrestre en utilisant l'algorithme basé sur les segments 3D.
- Recalage de l'image aérienne et l'image terrestre en utilisant l'algorithme basé sur les segments 3D.
- Recalage de l'image aérienne et le LiDAR terrestre en utilisant l'algorithme basé sur les segments 3D.

Dans le cadre de cette thèse, nous avons choisi de traiter le problème du recalage aérien/terrestre sous forme de problème du recalage de l'image aérienne et l'image terrestre. Ce choix n'est pas arbitraire, il est basé sur les raisons suivantes :

- Nous avons déjà étudié le recalage (données LiDAR/données LiDAR) où notre algorithme basé sur les polygones planaires a prouvé son efficacité pour traiter le problème du recalage intérieur/extérieur.
- Nous avons déjà étudié le recalage (image/ données LiDAR) comme un deuxième objectif de notre thèse où notre algorithme basé sur les segments 3D a prouvé sa performance.
- Comme l'algorithme basé sur les segments 3D a prouvé son efficacité pour résoudre le problème du recalage (image/données LiDAR) et recalage (données LiDAR/données LiDAR), recalage basé sur les ouvertures, nous sommes motivés pour tester sa performance pour résoudre des problèmes de recalage image/im-

age.

Nous avons réalisé le recalage d'une ortho-image aérienne et une séquence d'images terrestres. Ce recalage a été realisé en utilisant l'algorithme basé sur les segments 3D après avoir appliquer les adaptations suivantes :

- **Suppression du cluster vertical pour les données terrestres :**

  Comme pour les données aériennes, la plupart des lignes détectées se trouvent sur le sol (plan horizontal), nous avons donc supprimé le cluster vertical des données terrestres. Puis nous avons associé n'importe quelles paires de clusters terrestres avec n'importe quelles paires de clusters aériens s'ils sont un angle compatible.

- **Sélection de trois paires de segments pour RANSAC :**

  Pour l'algorithme original, à chaque itération de RANSAC, nous sélectionnons aléatoirement une association valide de cluster, puis nous sélectionnons aléatoirement un segment 3D dans chacun des clusters associés. Pour les données aériennes, nous avons choisi de travailler avec une ortho-image, où tous les segments sont coplanaires. Dans cette situation, si nous n'utilisons que deux paires de segments, nous obtenons une estimation d'échelle dégénérée. En ajoutant une troisième paire de segments, nous pouvons obtenir des informations supplémentaires sur la distance entre les segments, ce qui peut aider à estimer correctement le facteur d'échelle.

Les différentes solutions proposées dans le cadre de notre thèse ont dépassé les limites des algorithmes existants et ont prouvé leur efficacité pour résoudre les problèmes de recalage rencontrés par le projet ANR BIOM.

Mots clés

Recalage, Primitives géométriques, Polygones planaires, Segments 3D, Ouvertures, Energie globale robuste, Optimisation, RANSAC.

I would like to dedicate my thesis to my beloved parents.

Acknowledgements

This thesis has been a rich and unique experience for me that cannot be completed without thanking the people who have supervised, helped, and supported me.

I would like to thank my supervisors, Prof. Pascal MONASSE and Dr. Bruno VALLET, for the trust they put on me to conduct this challenging research work , their constant support and guidance, their precious advice and their availability throughout this thesis.

I also warmly thank the members of my jury, Prof. Beatriz MARCOTEGUI, Dr. Tania LANDES, Dr. Raouf BEN JEMAA, Prof. Nicole VINCENT and Prof. Paul CHECCHIN, for accepting to evaluate my work and providing invaluable comments on the report and during my defense.

I also thank, and without exception, all the members of IMAGINE research group.

A big thank you to Julien Gaubil, Nermin Samet, Pierre-Alain Langlois, Thomas Belos and Narimane Madi.

I cannot finish without having a thought for my family. A big thank you for my family for their indefinite support and especially my parents, my brother Soufiene and my sister-in-law Hassna.

Chapter 1

Context and research problem

1.1 Introduction

Our thesis is part of the Building Indoor/Outdoor Modeling (BIOM) project that aims at automatic, simultaneous indoor and outdoor modelling of buildings from heterogeneous data. The heterogeneity is both in data type (image and Light Detection and Ranging (LiDAR)) and acquisition platform: terrestrial indoor/outdoor or aerial acquisition. The objective is to achieve a complete, geometrically accurate, semantically annotated but nonetheless lean 3D CAD representation of buildings and objects they contain in the form of a Building Information Modeling (BIM) that will help manage buildings in all their life cycle (renovation, simulation, deconstruction). The first issue of such modeling is thus to precisely register the data. The work carried out has confirmed that the environment and the type of data drive the choice of the registration algorithm. So, the objective of this thesis is to explore fundamental properties of the data and the acquisition platforms in order to propose potential solutions for all the registration problems encountered by the BIOM project.

1.2 BIOM project

The Building Indoor/ Outdoor Modelling (BIOM) project aims at automatic, simultaneous indoor and outdoor modelling of buildings from images and dense point

clouds. The goal of the project is to achieve a complete, geometrically accurate, semantically annotated but nonetheless lean 3D CAD representation of buildings and objects they contain in the form of a Building Information Model (BIM) that will help manage buildings in all their life cycle (renovation, simulation, deconstruction). We view indoor and outdoor building modelling as a joint process where both worlds fruitfully cooperate and benefit one another both in terms of semantics and geometry. The hope is that this holistic scene understanding and reconstruction approach will lead to more complete, correct, and geometrically accurate building models.

The first challenge of the BIOM project is to accommodate for heterogeneous data as full building modeling calls for data acquisition inside and outside the building but also from an aerial point of view to model roof. The BIOM project also aims at exploiting the complementarity of image and LiDAR data. Another challenge is coping with incomplete data due to occlusions by furniture inside and urban and mobile objects outside. Last but not least, BIOM aims at modeling a large variety of architectural styles, different interior scene layouts, and a high amount of different objects that may be contained within the scene. State-of-the-art approaches treat outdoor and indoor worlds separately: most indoor reconstruction approaches focus on detailed modelling of single rooms whereas only very few have dealt with 3D modelling of complete floors (under Manhattan world assumptions). To the best of our knowledge, no works have been proposed, yet, that model buildings outdoor and indoor simultaneously within one single comprehensive framework. The figure 1.1 shows the main acquisition modes of the BIOM project. This project is funded by ANR, the French national research funding agency.

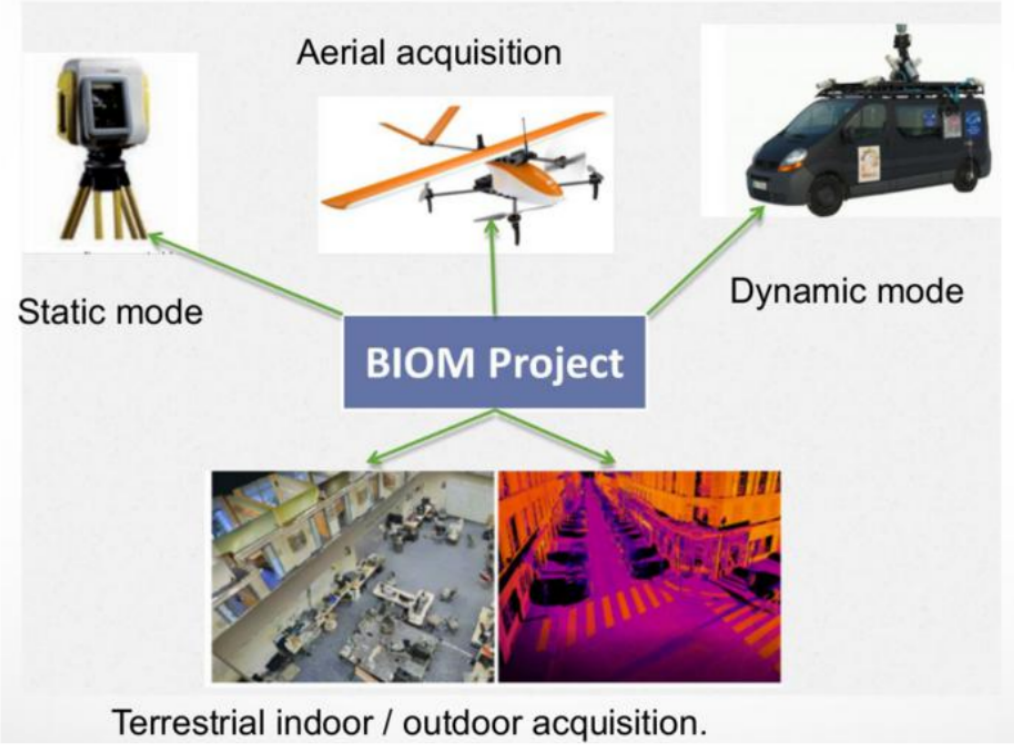


Figure 1.1: The main acquisition modes of the BIOM project

1.3 Objectives

The aim of this PhD is to address all registration issues faced by the BIOM project, as a major objective of the project is exploiting jointly data from various sensors (images/LiDAR) and viewpoints (terrestrial indoor/outdoor, aerial). Four key objectives for the BIOM project have been identified:

- Registration of outdoor heteregeneous data.
- Registration of indoor data.
- Indoor/outdoor landmark extraction.
- Global indoor/outdoor registration.
- Image/LiDAR registration.

1.4 Building Information Modeling

A Building Information Modeling (BIM) represents a comprehensive digital representation of a built facility with great information depth. It is based on parametric CAD

technology and represents a new approach of “Virtual Building Construction”(Woo, 2006). The basic principle of BIM is to share knowledge resource for information about a facility forming a reliable basis for decisions during its life-cycle; defined as existing from earliest conception to demolition (Borrmann et al., 2018). The collaboration by different stakeholders at different phases of the life cycle of a facility can help to insert, extract, update or modify information in the BIM to support and reflect the roles of that stakeholder.

BIM has a great potential for operation and maintenance (O&M) and facility management (FM) of buildings. However, most of the existing buildings have no BIM. Research in creating BIM for existing buildings has received growing attention in recent years. But, it remains a challenging task (Hossain and Yeoh, 2018)

1.5 Registration problem

The principle of registration according to (Monnier et al., 2013) consists in making at least two sets of data geometrically consistent. Generally, one of the data sets serves as a reference on which the second is registered. The goal is to determine the transformation necessary to best reconcile the data with each other. For this, we can summarize a registration problem in two important steps. The first consists in extracting common features in the datasets. The second consists in using these features in order to determine the optimal transformation to apply.

1.6 Images registration

Registration is a fundamental task in image processing used for aligning two or more images of the same scene with reference to a particular image. These images can be taken at different times, from different sensors, or from different viewpoints. Image registration became a large research axis in medical sciences, remote sensing and in computer vision.

1.6.1 The main image registration steps

According to (Nag, 2017), the following steps are necessary to solve an image registration problem

1. **Feature detection:**

   It is the key task of the Image Registration process which can be realised automatically or manually. These features must be easily detectable and physically interpretable and identifiable.

2. **Feature matching:**

   This step essentially establishes the correspondence between the features detected in the images that we want to register them. Different feature descriptors and similarity measures besides spatial relationships among the features are adopted to set up an accurate accordance.

3. **Transformation estimation:**

   For alignment of the sensed image with the reference image the parameters of the geometric transformation must be estimated. These parameters are computed with the established feature correspondence obtained from the previous step.

1.6.2 Criteria of image registration techniques

A registration model can be characterized by the following elements:

1. **Dimensionality:**

   This specifies the dimensions of different possible registrations. It may be 2D-2D, 2D-3D or 3D-3D.

2. **Domain of transformation:**

   It can be global when the whole image must be registered or it can be local when only a part of the image is taken into consideration for the registration.

3. **Geometric transformation:**

   It may be rigid (translation, rotation, reflection), affine (translation, rotation, scaling, reflection, shearing), projective or non-linear.

4. **Attributes:**

   These are extracted features from images that can guide the registration process.

5. **Inter-image distance:**

   It is a function that can be used to measure a distance between the attributes of the two images to be registered. This distance must be minimum when the similarity between the two images is maximum.

6. **Optimization method:**

   The last fundamental aspect of a registration algorithm is the choice of the optimization method whose goal is to determine the best transformation that minimizes or maximizes a cost function. In optimization, the convexity of a criterion is a desirable property.

1.6.3 Features detection approaches

The choice of the registration features is an essential step which is made according to three approaches (Vincent et al., 2013).

Iconic approach

Iconic methods can be considered low-level approaches. Iconic methods use the dense information carried by the gray levels of the image.

Geometric approach

This approach consists in extracting common geometrical characteristics in the images, which are called primitives (e.g: points, curves). These units carry high level information. The choice of the primitives is based on certain criteria:

- Easy feature detection.

- Distribution of primitives over the entire image.
- Robustness to noise, artifacts, changes related to acquisition.

Hybrid approach

This method consists of combining several different types of attributes. We can distinguish three types of combination: the combination of geometric primitives of different natures, the combination of different information from gray levels and the combination of geometric and iconic approaches.

1.7 Point clouds registration

The need to digitally represent the world around us has led to the development of technologies to scan an environment in three dimensions. Scanning methods produce a set of 3D points called ”point cloud”. Point clouds has become the primary data format to represent the 3D world with the fast development of high precision sensors such as LiDAR. Because the sensors can only capture scans within their limited view range, we need several points of view to cover the scene of interest. The point clouds from these different points of view are each defined in a reference relative to the sensor. So a registration is required to generate a large 3D scene. The aim of registration is to grouping all point clouds in a common coordinate system. There are two main families of registration methods depending on the number N of input point clouds: pairwise ($N = 2$) and multi-view ($N > 2$) registration.

1.7.1 Pairwise registration

We can consider pairwise registration as a special case of multiview registration where only one pair of overlapped views is considered. According to (Gojcic et al., 2020), the traditional pairwise registration pipeline consists of two stages: the coarse alignment stage, which provides the initial estimate of the relative transformation parameters and the refinement stage that iteratively refines the transformation parameters by minimizing the 3D registration error under the assumption of rigid transformation.

1.7.2 Multiview registration

Multiview registration can be formulated as an optimization problem based on incorporating cues from multiple views. The goal of optimization is to find a transformation for each view such that the overlapping regions of the transformed views are aligned with each other as closely as possible. This kind of methods aim at resolving hard or ambiguous cases that arise in pairwise methods. If the transformations are restricted to be rigid, the problem is characterized as multiview rigid registration. Multiview registration is more complicated than pairwise registration. The simple integration of local pairwise alignments will lead to the loop closure problem. Thus, a global method should be adopted. The process of solving the loop closure problem is also called loop- closing. Certain global methods take advantage of local pairwise alignments and perform loop-closing by diffusing the transformation errors in the initially aligned view pairs (Tang and Feng, 2015).

1.8 Contributions

The work carried out has confirmed that the environment and the type of data drive the choice of the registration algorithm.

Our contribution consists in exploring the fundamental properties of the data and acquisition platforms in order to propose primitives based registration algorithms able to solve all registration problems encountered by the BIOM project.

The basic idea of these algorithms consists in the definition of a global robust energy between the extracted primitives from the data-sets to register and the proposal of a robust method for optimizing this energy based on the RANSAC paradigm. In a building environment, most objects are composed of planar surfaces delimited by straight lines. In addition, we can represent a building as a set of façades, where the openings are the most obvious common elements of these entities. So, our selected primitives for this work are: planar polygons, 3D segments and openings (vertical rectangles).

To extract these primitives:

- We have proposed an efficient extraction algorithm (openings detection).
- We have adapted existing methods (planar polygons detection).
- We have improved an existing method (3D segments detection and reconstruction from images).
- We have directly used an existing method (3D segments detection from LiDAR data).

Our contribution has two aspects: methodological aspect and application aspect as shown in Figure 1.2.

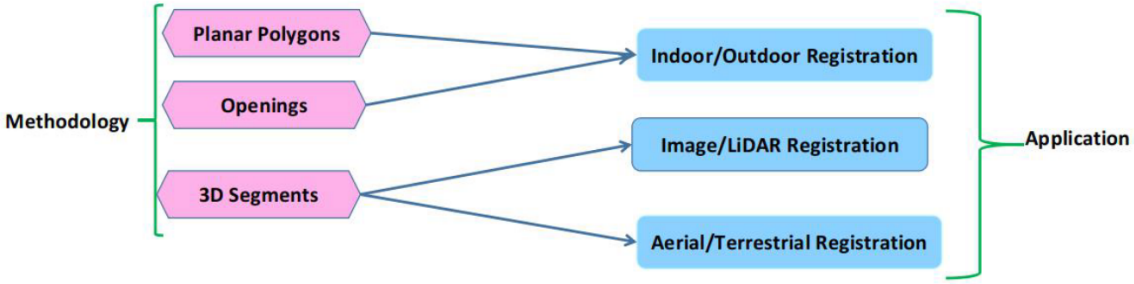


Figure 1.2: Overview of our contribution

Our methodological aspect consists in proposing registration methods based on our selected primitives:

- Planar polygons based registration algorithm.
- Openings based registration algorithm.
- Hybrid registration algorithm.
- 3D segments based registration algorithm.

Our application aspect consists in using the proposed algorithms to solve the registration problems encountered by the BIOM project. All the developed algorithms as part of our thesis follow one of the three possible indicated schemes in the Figure 1.3 to optimize the global robust energy. The choice of the schema to follow depends on the type of the geometric primitives (heterogeneous, homogeneous) extracted in each data-set as well as the quantity.

- If the two data-sets contain few homogeneous primitives, we can directly apply our RANSAC.
- If the two data-sets contain many homogeneous primitives, we can cluster them and associate the obtained clusters to simplify our RANSAC.
- If the two data-sets contain heterogeneous primitives, we must associate them in each data-set before applying our RANSAC.

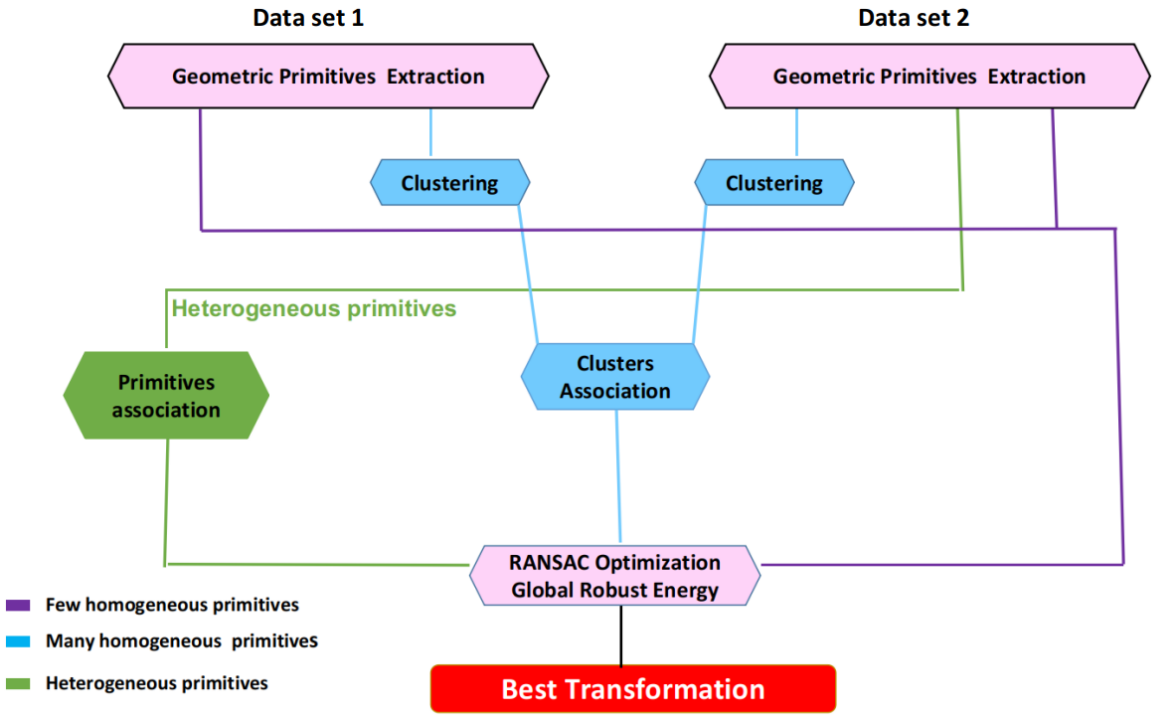


Figure 1.3: Possible schemes for our algorithms.

The main registration problems encountered by the BIOM project are as follows.

1.8.1 Indoor/outdoor registration problem

The registration of indoor and outdoor scans with a precision reaching the level of geometric noise represents a major challenge for Indoor/Outdoor building modeling. The lack of overlap between indoor and outdoor data is the most encountered obstacle, more so when both data sets are acquired separately and using different types of sensors. Among our proposed algorithms, we can consider the following algorithms as potential solutions of this challenging problem:

A planar polygons based algorithm

It is an efficient algorithm based on polygon detection and matching. Its strong points lie in the fact that it exploits with the very small overlap between indoor and outdoor scans of the same building by extracting points lying in the buildings' interior from the outdoor scans as points where the laser ray crosses detected façades. It follows the RANSAC paradigm and uses a global robust energy in order to select the best solution.

An openings based algorithm

This proposed algorithm is based on openings detection and matching. An opening is considered among the common entities that can be seen from inside and outside. It is a rectangular shape composed of four segments. The algorithm registers the detected openings from both inside and outside by minimizing a global robust energy between the corresponding segment sets. This minimization is based on the RANSAC paradigm.

Hybrid registration algorithm

The two previous algorithms are efficient to deal with indoor/outdoor registration problem, but they have some limitations. In order to exceed these limitations, we have combined them. This combination has produced a very efficient hybrid algorithm. This new algorithm has proven its performance to solve the indoor/outdoor registration problem.

1.8.2 Image/LiDAR data registration problem

Image/LiDAR data registration is a very difficult problem given the very strong difference in modality. This difference makes it difficult to extract comparable features between the two modalities. Conversely, with our primitives and our RANSAC based approach, we assume that the best transformation is the one which optimizes the global robust energy between all the extracted primitives and therefore which is the most consistent with all the information of the datasets to be registered. Our 3D segments based registration algorithm represents a potential solution for this regis-

tration problem. This algorithm based on the definition of a global robust energy between two segments sets and the minimization of this energy according to the RANSAC paradigm.

1.8.3 Aerial/Terrestrial registration problem

The aim of our thesis is not only registration of heterogeneous data (image and LiDAR data) but also the registration of data acquired by heterogeneous acquisition platforms (aerial/terrestrial). By exploring our previous solutions, we can deal with this problem in the following forms:

- Aerial LiDAR/ Terrestrial LiDAR registration: using planar polygons based algorithm
- Aerial image/ Terrestrial LiDAR registration: using 3D segments based algorithm
- Aerial LiDAR/ Terrestrial image registration: using 3D segments based algorithm
- Aerial image/ Terrestrial image registration: using 3D segments based algorithm

For the aerial/terrestrial registration problem, we chose to focus on the particular problem of aerial image / terrestrial image registration for the following reasons:

- we have already studied the LiDAR data /LiDAR data registration in the first part of our thesis (chapter 4).
- we have already studied the image/LiDAR data registration in the second part of our thesis (chapter 5).
- As the 3D segments based algorithm has proven its effectiveness in solving the problem of image/LiDAR data registration and LiDAR data /LiDAR data registration (openings based registration), we hoped to test it in solving the image/image registration problem.

We have adapted our 3d segments based algorithm to solve this problem.

1.9 Thesis outline

This thesis is organized as follows:

- **Chapter 2: Data description**

  This chapter is dedicated to the description of the data used for this study (image/LiDAR data) and the acquisition platforms.

- **Chapter 3: Primitives detection**

  This chapter is divided into two main parts: A first part is dedicated to the classification of primitive extraction methods, followed by an explanation of the criteria for evaluating the performance of algorithms. A second part deals with the methods of extraction of the geometric primitives which interest us, namely the planar polygons, the 3D line segments and the openings.

- **Chapter 4: Indoor/Outdoor Registration**

  This chapter deals with the indoor/outdoor registration problem. We start with a state of the art followed by the proposal of two initial solutions (a planar polygon based solution and an openings based solution. Then, we combine these two solutions to overcome their limitations.

- **Chapter 5: Image/LiDAR data Registration**

  This chapter deals with the problem of heterogeneous data (Image/LiDAR data) registration. We start by classifying some existing methods according to their types and attributes, citing the advantages and limitations of each category. Then, we explain our 3D segments based algorithm.

- **Chapter 6: Aerial/Terrestrial registration**

  This chapter has been approached the problem with the registration of data acquired by heterogeneous acquisition platforms (aerial/terrestrial). We start by illustrating some existing works followed by the adaptation of the 3D segments based algorithm to solve this kind of problem.

- **Chapter 7: General Conclusion**

  This chapter reflects on the contributions of the thesis and suggests on directions of future works.

1.10 Publication List

Category	Title	Status	Corresponding chapter
Conference	Towards efficient indoor/outdoor registration using planar polygons	Published (ISPRS conference 2021) + **ISPRS Best Young Author Award**	Chapter 4
Conference	Detecting openings for indoor/outdoor registration	Published (ISPRS conference 2022)	Chapter 4
Conference	A 3D segments based algorithm for heterogeneous data registration	Published (ISPRS conference 2022) + **Best Poster Paper Award**	Chapters 5, 6
Journal	Primitives based algorithm for hybrid registration	Planned (ISPRS journal)	All
Journal	MLSD performance improvement	Planned (IPOL journal)	Appendix B

Table 1.1: Publication List

Chapter 2

Data description

2.1 Introduction

Generally, to solve any problem, you have to choose the appropriate algorithm to apply. This choice should not be random. It must be guided by the type of data and the environment. The aim of our thesis is the development of potential algorithms for heterogeneous data registration in building environment. The heterogeneity is both in data type (image and LiDAR) and acquisition platform: terrestrial indoor/outdoor or aerial acquisitions. To achieve this objective, we must understand and study the fundamental properties of this data, the acquisition platforms and the environment.

2.2 Image data

2.2.1 Image definition

There is no universal definition of an image. According to (Mishra et al., 2017), an image is a 2D function $f(x, y)$ where (x, y) is the coordinate in two dimensional space called a pixel, which represents the smallest unit of the image , and f is the color at that coordinate. A digital image is a rectangular array of pixels also called a Bitmap. Pixel is an abbreviation of picture element. Each pixel stores a color at this point in the image. In general, digital images have two types of color representation:

- Binary images assign only one of two values to a single pixel: 1 or 0 (black or white) and are used to represent geometric sets or image masks.
- Black and white images, also called grayscale or panchomatic images are made of different shades of gray corresponding to the luminosity received by the sensor on the full visible spectrum (400 to 800nm). These different shades are usually sampled on 8 bits between 0 (black) and 255 (white) and intermediate values correspond to different shades of gray.
- Color images are made up of colored pixels. The spectrum – the band of colors produced when sunlight passes through a prism – includes all primary colors, of which the human eye can perceive seven to ten million. The electronic capture and display of color is commonly based on the RGB (Red, Green, and Blue) color system which represents each channel on 8 bits (from 0 to 255) leading to a 24-bit representation that defines 16.8 million colors.

2.2.2 Image acquisition model

The images are generated by combination of a light source and the reflection or absorption of the energy by the elements of the scene of interest. In a general way, to capture the image, we use a sensor according to the nature of the light. The light energy is transformed into digital image using the sensor. By combining the input electrical energy and sensor material, that is responsive to the particular energy that is being detected, incoming illumination energy is transformed into voltage (Mishra et al., 2017). The output waveform is the response of the sensor and this response is digitized to obtain a digital image as shown in Figure 2.1.

2.2.3 Techniques to perform image acquisition

Image Acquisition process totally depends on the hardware system. An image sensor converts an image into a digital signal. Nowadays, there are two main technologies which is used in image sensor: Complementary Metal Oxide Semiconductor (CMOS) and Charge Coupled Device (CCD).

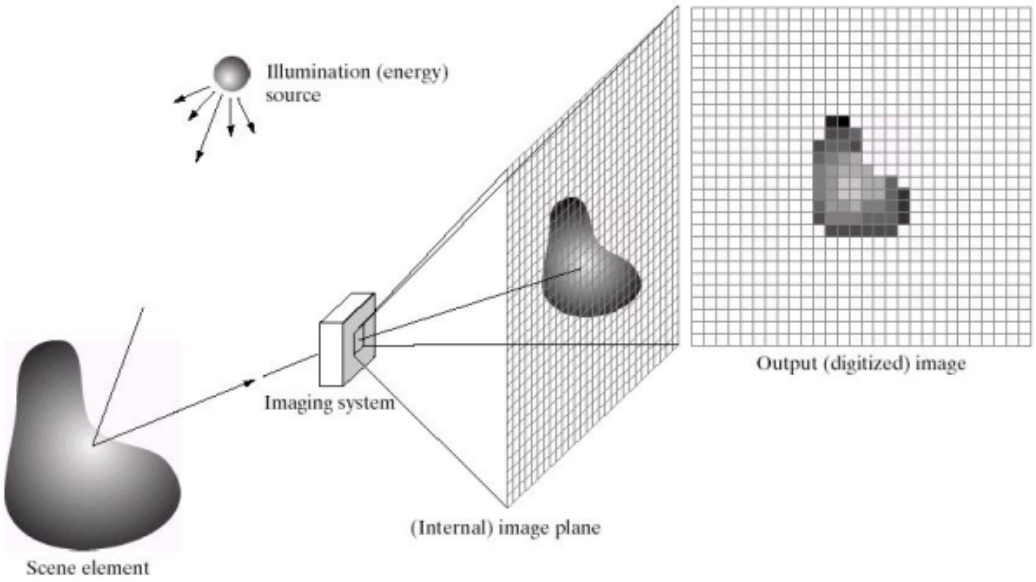


Figure 2.1: Image Acquisition Model(Mishra et al., 2017)

CMOS

According to (Mehta et al., 2015), the working principle of CMOS sensor is the following: photons from objects are falls on the photodiode and it converts it into the charges. Then this charge will apply to the capacitor that will convert the charges into the voltages. Finally *Analog to digital converter* will convert the voltages into the digital signal as shown in Figure 2.2. There are two CMOS detector types: *Passive Pixel* and *Active Pixel*.

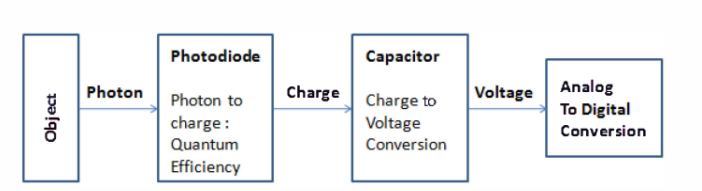


Figure 2.2: Image Acquisition Model (Mehta et al., 2015)

CCD

According to (Mishra et al., 2017), CCD image sensors convert light into electrons . Once the light is converted into electrons, it reads the value (accumulated charge) of each cell in the image. A CCD transports the charge across the chip and reads it at one corner of the array. An analog-to- digital converter (ADC) then turns each pixel's value into a digital value by measuring the amount of charge at each photosite and

converting that measurement to binary form.

The Table 2.1 studies the advantages and limitations of the two types of image sensors. This study is taken from (Mishra et al., 2017)

image sensor	Advantages	Limitations
CMOS	+ Low Cost + Consume Less Power + High Speed	- Low fill factor - Low light sensitivity - Low Charge Capacity - Pixel Uniformity and Noise
CCD	+ High Quality Image + High Quantum Efficiency + Low dark current + Very low noise	- Highly nonprogrammable - High power Consumption

Table 2.1: Comparison between CCD and CMOS

Overall, CMOS sensors are much less expensive to manufacture than CCD sensors, are rapidly improving in performance. They traditionally consume little power. CCD sensors consume as much as 100 times more power than an equivalent CMOS sensor. CCD sensors have been mass produced for a longer period of time, so they are more mature. They tend to have higher quality pixels, and more of them (Mishra et al., 2017).

2.3 LiDAR data

2.3.1 LiDAR concept

LiDAR is an acronym for light detection and ranging. It refers to a remote sensing technology that emits intense, focused beams of light and measures the time it takes for the reflections to be detected by the sensor. This information is used to compute ranges, or distances, to objects. In this manner, LiDAR is analogous to radar (radio detecting and ranging), except that it is based on discrete pulses of laser light (Carter et al., 2012). As LiDAR uses its own laser light energy source to illuminate an area of the atmosphere or the targets to be studied, it is considered an active optical analysis method (Flamant, 2019).

2.3.2 LIDAR principle

According to (Mehendale and Neoge, 2020), LiDAR employs the method of sending laser light on to the target and measuring the reflected light back to recognize the variation in wavelength and arrival time of the reflected light. The measurement is done by calculating the laser return times and their wavelengths. It generates precise, high quality, and sometimes even a three-dimensional map of the environment that it scans. The generated map of the area in focus, helps in its characterization and examination. A typical LiDAR system consists of a scanner, laser, and sometimes specialized GPS receiver as well. Other elements that are essential for data collection and analysis are optics and photodetectors.

The design principle behind LiDAR is the reflection of light. This principle is to shine a light beam on to a surface and calculate the time it takes to return to its source. The LiDAR system sends laser light on to the target and measures the reflected light to see the variation in wavelength and arrival time of the reflected light. From these measurements, it can calculate the distance to draw the digital representation of the target. Since light travels at a very high speed, the calculation of the exact distance through LiDAR is very fast. The formula that the analysts use to calculate the distance is given in equation (2.1)

$$D = c\left(\frac{\Delta T}{2}\right) \tag{2.1}$$

where, D = The distance of the object.

c = Speed of light.

Δ = Time required by the light to travel.

The LiDAR system fires many laser lights on to the surface. The sensor on the system measures the time taken for the reflected light to reach the sensor. This goes on repeating until a complex map of the surface is constructed.

2.3.3 LiDAR acquisition platforms

LiDAR technology can perform in two forms: static (Terrestrial laser scanning) and dynamic (Mobile laser scanning, Mobile mapping system, Airborne Laser Scanning and Unmanned Aerial System).

Terrestrial laser scanning

Terrestrial laser scanning (TLS), also referred to as terrestrial LiDAR (light detection and ranging) or topographic LiDAR, is a technique which uses laser light in order to measure with high speed in a dense regular pattern directly 3D coordinates of points on surfaces and surface brightness from terrestrial position (Pfeifer, 2007). This measurement is realized by emitting laser pulses toward the target and measuring the distance between it and the device (Dumbrell et al., 2019). The number and variety of applications of TLS instruments continue to increase.

By emitting laser pulses toward numerous points on land, TLS can acquire their XYZ coordinates. This acquisition is based on measuring the distance from the device to the target. The table 2.2 presents the main characteristics and performances of terrestrial laser scanners.

	Terrestrial laser scanning (TLS)
Absolute accuracy	1 cm at distance 100 m
Range	From 100 m to 6 km
Acquisition time	Variable according to the monitored surface (15 mn for a standard application)
Acquisition frequency	$\leq$ 300 kHz
Wavelength	0.78 μm to 1 mm
Field of vision	360° horizontally and up to 80° vertically
Equipment weight	Between 20 and 45 kg

Table 2.2: Main characteristics and performance of terrestrial scanner solutions (Coccia, 2021)

Mobile laser scanning

Mobile mapping system (MMS): According to (Kukko et al., 2013), we can consider a mobile mapping system as a kinematic surveying system utilizing motion tracking and time synchronized data acquisition, regardless of the platform in use. A GNSS receiver(s) with appropriate antenna(s) for positioning and an IMU for platform and sensor attitude determination and high-frequency positioning represent the main components of the navigation system.

Mobile laser scanning (MLS): Mobile laser scanning (MLS), is a rapid and flexible method for acquiring three-dimensional topographic data. Mobile Laser Scanning,

being a sub-category of MMS, based on Light Detection and Ranging (lidar) sensors is a rapid and flexible technology for acquiring high-resolution three-dimensional data. With state-of-the-art sensors, the achieved point clouds capture object details with good accuracy and precision (Kukko et al., 2012). In general, the spatial coverage of data in MLS is achieved by the movement of the vehicle and motion-tracking positioning devices, as shown in Figure 2.3. The survey is conducted as the ground vehicle moves around, while the navigation system tracks the vehicle's trajectory and sensor attitude, used to produce a 3D point cloud from the range data collected by the onboard scanners. When the ground vehicle moves around, the survey is conducted. The vehicle's trajectory and sensor attitude, used to produce a 3D point cloud from the range data collected by the onboard scanners, are tracked by the navigation system. The survey is conducted as the ground vehicle moves around (Kukko et al., 2013).

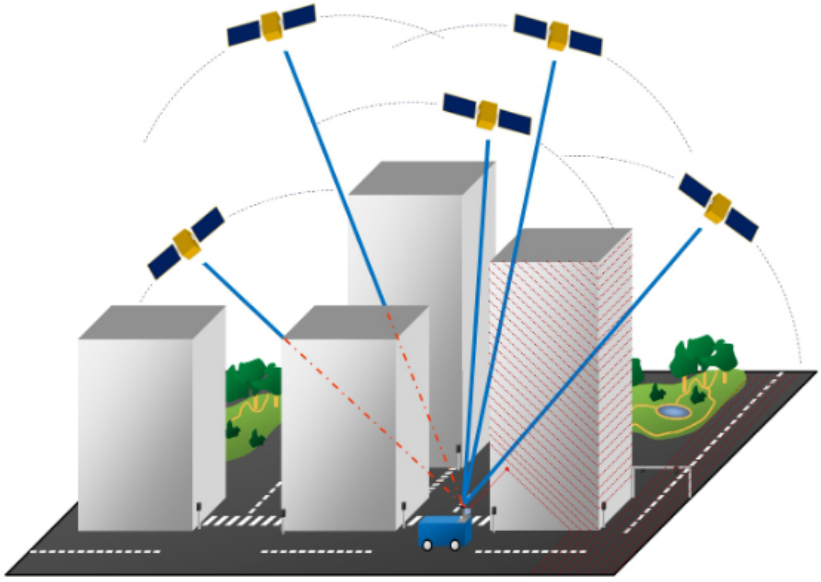


Figure 2.3: Mobile laser scanning utilizes GNSS/IMU positioning for direct georeferencing of point clouds (Kukko et al., 2013)

Sensor Topology: The sensors used to capture point clouds often have an inherent topology. According to (Guinard and Vallet, 2018), Mobile Laser Scanners sample a regular grid in (θ,t) where(θ) is the rotation angle of the laser beam and (t) the instant of acquisition. As the vehicle moves at a variable speed and can turn, the sampling is not completely uniform in space. For a 2π rotation in θ the pulses number N_p is not an integer, so it considers that a pulse of P_i has six neighbors as shown in

Figure 2.4. However, this topology concerns emitted pulses, not recorded echoes. One pulse might have 0 echo (no target hit) or up to 8 as most modern scanners can record multiple echoes for one pulse if the laser beam intersected several targets. They approached the problem by connecting an echo to each of the echoes of its pulses' neighbors as shown in Figure 2.5. This idea is a potential solution to keep all possible edge hypotheses before filtering them.

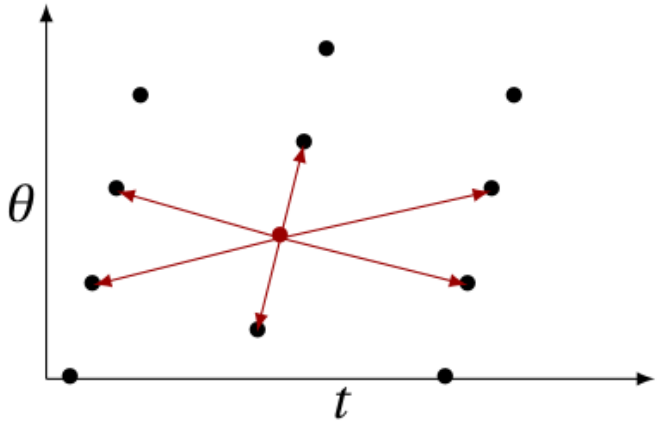


Figure 2.4: The pulse sensor topology forms a 6-neighborhood: the points considered is colored in red, and connection is denoted by a red arrow

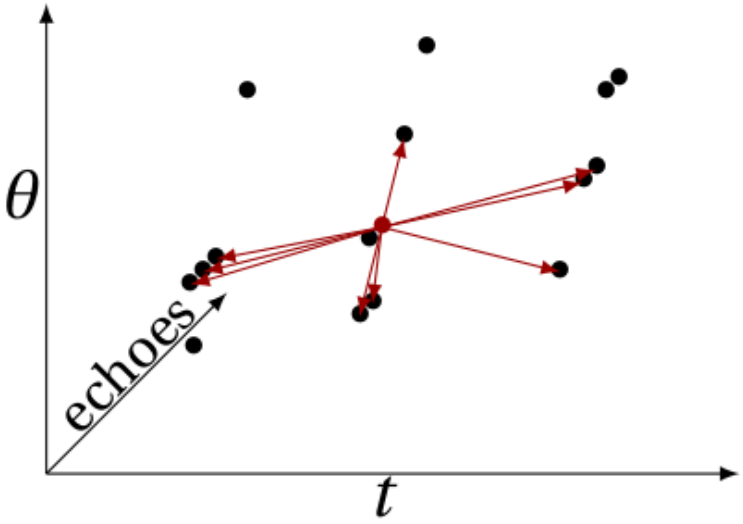


Figure 2.5: Echo sensor topology: each echo is connected to all echoes of all neighboring pulse, the points considered is colored in red, and connection is denoted by a red arrow

Airborne Laser Scanning

Airborne Laser Scanning (ALS) or Airborne Laser Swath Mapping (ALSM) is an active remote sensing technique, which uses laser scanning to record the surface of the

earth. It offers a range of opportunities for mapping and change detection. ALS produces high-resolution topographic data and allows very precise three-dimensional mapping of the surface of the earth (Mlekuž, 2018). With airborne laser scanning, it is possible to determine the plane-ground distance and the inclination angles of the scanning platform. These tasks can be performed by frequent sending of short laser pulses from a flying platform (plane) towards the ground and measuring the return times of the reflected signal (Kurczyński, 2019). The trajectory of the plane can be measured using both GPS and inertial measurement.

The Figure 2.6 shows a detailed illustration of ALS data acquisition.

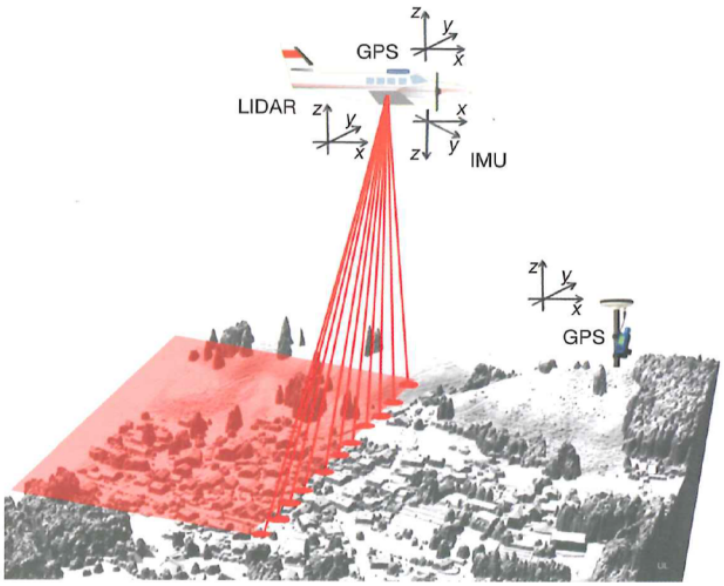


Figure 2.6: Schematical illustration of ALS data acquisition (Roncat, 2016)

Unmanned Aerial System

New ways to perform laser scanning surveys more cost-effectively are offered by Unmanned Aerial Systems (UASs), which opens doors to many new change-detection applications. According to (Kukko et al., 2016), UASs can generally be categorised into two types:

- Fixed-wing systems provide users with a longer operation time and support larger payloads due to better fuel economy. They allow for more speed, which makes this type of UAS more suitable for large areas or long-distance missions. But They are less favourable for small-area surveys with complex terrain or objects rich in features that need to be captured.

- Rotorcrafts allow for slow or even stationary flight speeds and offer excellent manoeuvrability. They do not require an airstrip for take-off and landing. Therefore, they enable the use of relatively low-cost sensors.

Table 2.3 shows a simple comparison between the main laser scanning systems: MLS, TLS and ALS.

Platforms	System Abbreviation	Scanning Perspective	Scanning Range	Point Cloud Density	Application Areas example
Airborne	ALS	Top view	Surface shape	Relatively sparse	Terrain mapping, 3D urban areas
Vehicle	MLS	Side view	Stripe shape	Dense	Road mapping, 3D urban areas
Tripod	TLS	Side view	Point shape	Dense	Deformation monitoring, reverse engineering

Table 2.3: Comparison of LiDAR systems mounted on different platforms (Cheng, Chen, Liu, Xu, Wu, Li and Chen, 2018)

2.4 Conclusion

Data are the basic and input elements of any processing chain. They guide the choices of the algorithms to be applied in order to solve the studied problem. In this chapter, we have focused our study on the fundamental properties of the data used in the context of our thesis (image and LiDAR). This study is essential to defend the choice of the proposed solutions in the next chapters.

Chapter 3

Primitives detection

3.1 Introduction

3D data can be represented in several forms. The most used representations are: depth map and point cloud. Each representation is adapted to some families of processing methods. The core of 3D data processing is the extraction and the reconstruction of geometric information. The point is considered a basic element of geometry. By grouping a set of points, we can create a more complex element which represents a geometric primitive. In a building environment, most objects are composed of planar surfaces delimited by straight lines. In addition, we can represent a building as a set of facades, where the openings are the most obvious common elements of these entities. So, we must effectively and precisely extract these components to be able to use them as attributes for more complex data processing such as registration and modeling. This chapter is divided into two main parts: A first part is dedicated to the classification of primitive extraction methods, followed by an explanation of the criteria for evaluating the performance of algorithms. A second part deals with the methods of extraction of the geometric units which interest us, namely the planar polygons, the lines, and the openings.

3.2 Classification of primitives extraction methods

Several methods have been proposed in the literature to deal with the problem of detecting primitives in 3D point cloud. We can classify these methods in four categories.

3.2.1 Clustering based methods

The clustering is a way for extracting geometric primitives from a 3D point cloud. This kind of algorithms is in the category of unsupervised learning and requires in most cases a post-processing step to refine the results. The most used algorithms for this kind of methods are:

- **Region growing (RG):**

  It consists in segmenting the 3D data using neighborhood information. It groups iteratively nearby points until some criterion is reached. The proximity may be based not only on spatial distance but also on similarity between local features (Hojjatoleslami and Kittler, 1998).

- **K-Means:**

  It consists in fixing a number k of clusters and randomly selecting k points as seeds. Points of the cloud are grouped according to the seed they are closest to based on some distance metric. These seeds move to the centroid of their group that represent the centroid of the detected clusters. The process iterates the steps assignment/motion until nothing changes (Yadav and Sharma, 2013).

- **J-Linkage:**

  This method deals with the problem of fitting multiple instances of a model in the presence of noise and outliers. The characteristic function of the set of random models that fit the point is associated to each point. Then, points having a similar characteristic function are clustered using a tailored agglomerative clustering (J-linkage). The method does not require adjustment of the parameters except the consensus threshold (Toldo and Fusiello, 2008).

3.2.2 Vote accumulation based methods

This family of method works in parameter space and needs a voting step in most cases. The Hough Transform (HT) has been proposed in (Duda and Hart, 1972) and it represents the standard of the accumulation algorithms. It constructs an accumulation space on a parameter sampling in which similar geometric elements coincide. The object candidates are obtained as local maxima in the accumulation space.

3.2.3 Hypothesis and validation based methods

This family of methods consists in the generation of primitive guesses and the selection of the best one using a validation criterion (cost function, energy...). The random sample consensus (RANSAC) is the most representative of this family of methods. It is a simple and powerful tool for 3D data segmentation. It is known to be particularly robust to outliers, provided sufficient samples are drawn. RANSAC was proposed in (Fischler and Bolles, 1981). The algorithm is iterative and based on a random selection of a minimal or small subset of samples, which is used for estimating the parameters of a model fitting them. Then, using its proper loss function (originally the number of outliers of the model), it obtains a score for the tested model, i.e., how well the model fits to the data. Several variants of the RANSAC paradigm have been proposed in order to improve its performance. (Torr and Zisserman, 2000) proposed two modifications of RANSAC called: MSAC (M-estimator SAmple and Consensus) and MLESAC (Maximum Likelihood Estimation SAmple and Consensus). The difference between RANSAC and these generalizations lies in the cost function.

3.2.4 Selection and decision based methods

Most detectors encounter multiple challenges such as: parameter tuning, control of false detection and execution time. The *a contrario* approach proposed in (Moisan et al., 2008) represents a potential solution for this kind of problem. A primitive detection algorithm has two main parts: model candidates selection and validation. Performing a validation based on the approach allows a control of the expected number of false detection, which does not require parameter settings. LSD (Line Segment

Method	Advantages	Limitations
Clustering	+ Meaningful segmentation + Few parameters + Special case: J-Linkage allows to detect multiple instances of a model at the same time	- Sensitive to initials conditions (seeds) - Sensitive to noise and outliers
Vote accumulation	+ Potential solution for missing data + Pretty natural for the detection of multiple objects	- Dependent on parameters space quantization
Hypothesis and validation	+ Simple + Robust to outliers	- Need many parameters - Requires a significant number of iterations depending on the data size
Selection and decision	+Potential solution for controlling false detection	- Dependent of fixed threshold over the probability of false alarm

Table 3.1: Comparison between primitives extraction methods

Detector) is the most popular algorithm of this family of methods. It has been proposed in (Von Gioi et al., 2012) to detect locally straight contours in an image.
A comparison between these four classes of methods has been carried out in Table 3.1, which underlines the advantages and limitations of each class of methods.

3.3 Performance evaluation

The most used metrics to evaluate the performance of a primitive detection algorithm are: the accuracy, timing and robustness.

3.3.1 Accuracy

The accuracy of the detection is obtained by comparing ground truth with the obtained results. If the error between the ground truth and the obtained result increases, the accuracy decreases and vice versa.

3.3.2 Robustness

A good algorithm must be robust to noise, incomplete data and outliers.

- **Robustness to noise:**

  The robustness to noise can be measured by the stability of the detection accuracy if the dataset contains noise.

- **Robustness of incomplete data:**

  For some acquisition tools, the output data may have missing parts. So the robustness of the algorithm is represented by its capacity to complete the shape if some parts of it are not captured, such as polygons with holes.

- **Robustness to outliers:**

  The robustness to outliers can be measured as the ability of the algorithm to distinguish between points coherent with the model (inliers) from the others (outliers).

3.3.3 Timing

This criterion is related to the speed of the algorithm execution and it depends on the data size and computing power of the devices. The data size is the representative size of the datasets targeted by the algorithm. We will only consider algorithms that process the input data in less than one minute.

3.4 Planar polygons extraction

3D data can be represented as a 3d point cloud. This point cloud can be unorganized (e.g., LiDAR data) or organized (e.g., range images). The high computation and memory costs of dense point cloud processing requires us to go through a concise and meaningful abstraction of the data. In man-made environments where most objects are represented by planar surfaces, the planar polygons are an adequate abstraction of the data.

3.4.1 State of the art

Modeling a 3D point cloud as a set of planes is known as the plane fitting problem. Plane fitting is a very active field of research. Hough transform, random sample

consensus (RANSAC) and region growing are the most popular approaches to detect planes in point clouds. The RANSAC algorithm (Fischler and Bolles, 1981) represents a potential solution to iteratively find an accurate model for observed data that may contain a large number of outliers and noise. An adaptation of RANSAC to the plane fitting problem has been proposed in (Schnabel et al., 2007). It ensures a precise and fast extraction of the plane provided that the parameters are properly adjusted. Hough transform(HT) (Borrmann et al., 2011) is also a robust method to detect plane. The high computation cost is its main limitation. Several improvements have been proposed to overcome this limitation. A new approach has been proposed in (Limberger and Oliveira, 2015). It works by casting votes for a set of clusters on a spherical accumulator using a trivariate Gaussian kernel. The clusters are generated by clustering sets of approximately co-planar points. An efficient triple-region growing method has been proposed in (Hu et al., 2020) for plane detection in rock mass point clouds. According to the bounding box and specified resolution, this method starts by dividing point clouds into small voxels. Afterwards, these voxels are regarded as growing units (seeds) and clustered into completed planes by region growing. Finally, the corresponding polygons are calculated by concave hull method. A new hypothesizing and selection strategy based framework has been proposed in (Nan and Wonka, 2017). The intersection of the extracted planar primitives has been performed in order to generate a reasonably large set of face candidates. Finally, an optimal subset of the candidate faces is selected through optimization. Polylidar3D is a concave polygon extraction method from 3D data which has been proposed in (Castagno and Atkins, 2020). This method perfoms by modelling the 3D point cloud as a polygonal (triangular) mesh. The obtained concave polygons represent flat surfaces in an environment with interior cutouts representing obstacles or holes.

The interest of polygons compared to planes are the following:

- The memory space required is considerably reduced (in terms of number of points).
- They enable the calculation of certain relevant features, for example, centroid, area, perimeter.

- The polygons have a spatial extent limited to areas where they have support points in input data. They therefore form a simple and compact summary of our scans.

3.4.2 Our proposed solutions

Due to its robustness to noise and outliers, Random Sample Consensus (RANSAC) has become the most popular method for LiDAR point cloud segmentation. Despite this success, it can generate false segments consisting of points from several nearly coplanar surfaces. False planes made up of points from different planes or roof surfaces represent a real obstacle for RANSAC (Xu et al., 2016). In order to overcome the limitations of RANSAC, we have exploited two methods depending on the nature of the data.

RANSAC Based on Sensor Topology

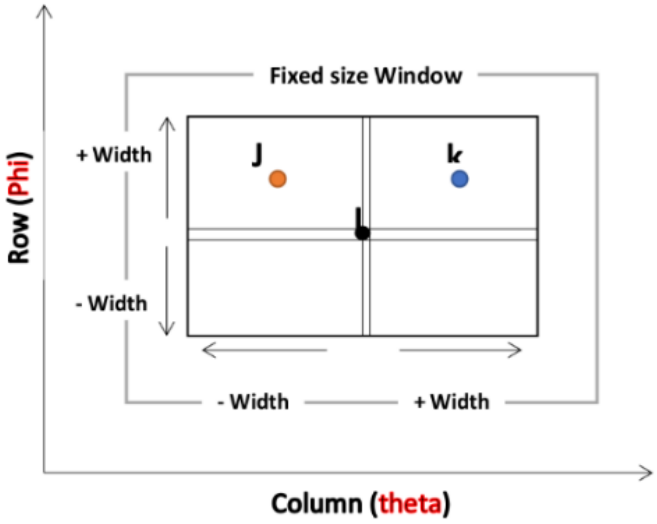


Figure 3.1: Schematics illustrating the sensor topology approach

For scans acquired by a Mobile Mapping System (MMS) (e.g., outdoor scans), we have access to the sensor topology (adjacency between successive pulses in the same line and between lines) so we can use it to enhance the polygon RANSAC detector. We use for that a recent method (Guinard et al., 2020) that exploits the sensor topology to extract compact planar patches instead of planes:

- Sample selection: as we are looking for compact planar patches, once a first sample point is drawn randomly from the point cloud, the other two are drawn

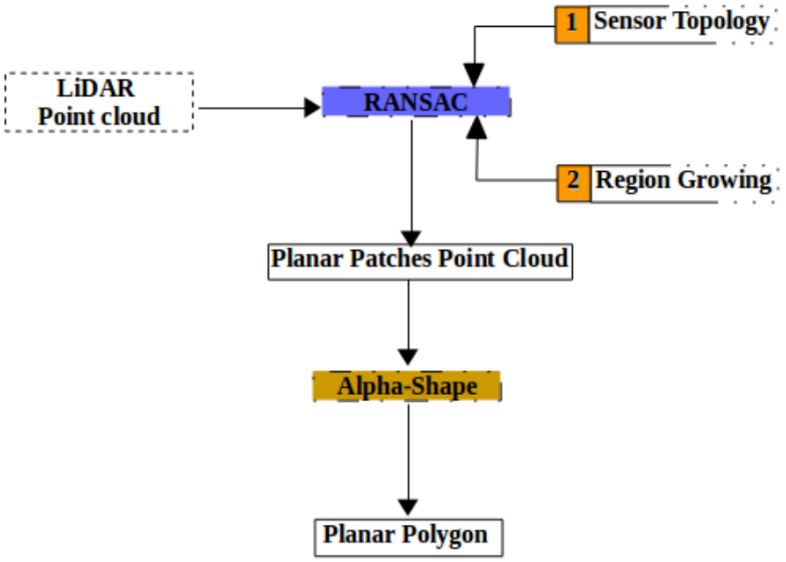


Figure 3.2: Pipeline details of sensor topology based RANSAC

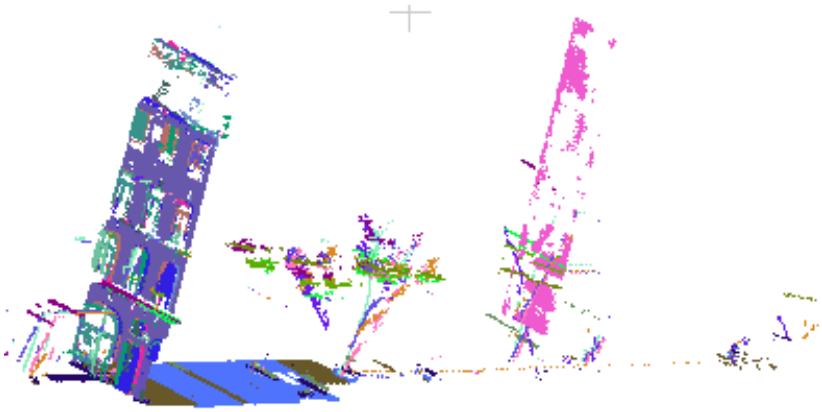


Figure 3.3: Inliers of the estimated planes from an outdoor scan computed with Sensor topology based RANSAC

in a local neighborhood (defined based on the sensor topology) as shown in Figure 3.1.

- Region growing: instead of computing the distances of all the points of the cloud to the hypothetical plane, a region is grown starting at the first sample in order to recover a compact planar patch.

At each iteration, the planar patch with most inliers is selected and approximated by a single polygon using the α-shape algorithm (Edelsbrunner et al., 1983). Figure 3.2 shows the pipeline steps. Figure 3.3 shows the inliers of the detected planes from an outdoor scan.

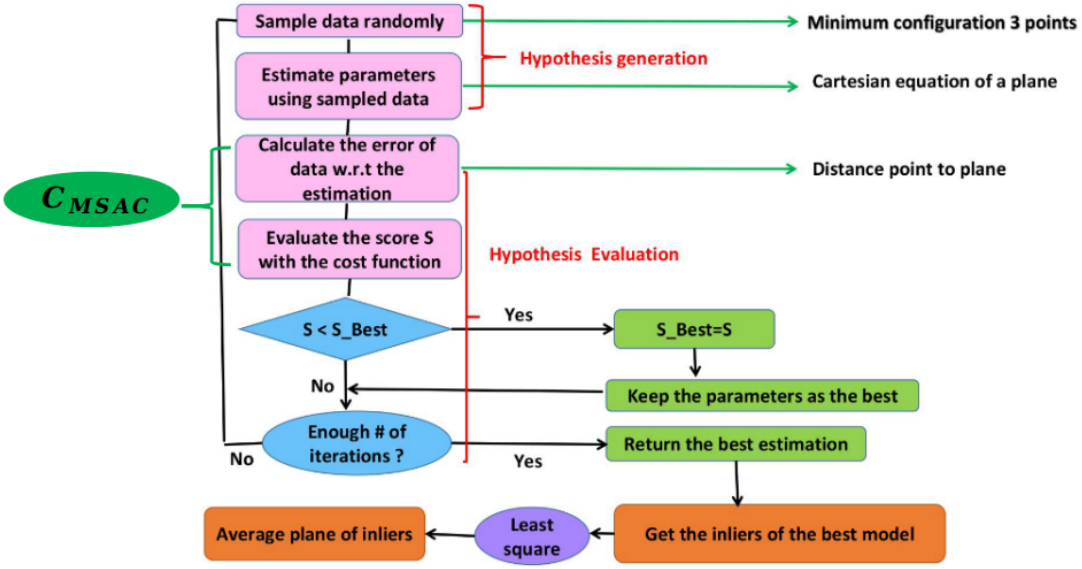


Figure 3.4: Robust estimation of a plane using MSAC.

Polygon MSAC

For scans acquired in a static mode (e.g., indoor scans), we do not have access to the sensor topology. So, we could not use the aforementioned method for the extraction of planar polygons. This is why, we have proposed a straightforward adaptation of M-estimator Sample Consensus (MSAC) which is a RANSAC extension that provides a potential solution to the spurious plane problem (Torr and Zisserman, 2000). Whereas RANSAC gives the same unit score to all inliers:

$$C_{\text{RANSAC}}(e_i) = \begin{cases} 0 & \text{if } \epsilon_i^2 < t^2 \\ 1 & \text{otherwise} \end{cases} \tag{3.1}$$

MSAC gives each point a penalty score measuring how well the point corresponds to the model:

$$C_{\text{MSAC}}(\epsilon_i) = \min(\epsilon_i^2, t^2) \tag{3.2}$$

where ϵ_i is the distance of LiDAR point P_i to the current hypothetical plane and t is a distance threshold, parameter of the algorithm.

As RANSAC, MSAC produces hypothetical planes by randomly selecting three input points. The score of the sample plane is simply the sum of (3.2) over all the points P_i of the input scan. When the scores have been computed for all planes, the one with the highest score is extracted, its inliers (points P_i such that $\epsilon_i < t$) are removed

from the point cloud, and the process is iterated until the score of the best plane is below a given threshold, as detailed in Figure 3.4. For each detected plane, we project its inliers to the plane and extract planar polygons by computing the $\alpha-$shape (Edelsbrunner et al., 1983) of these projected inliers. Figure 3.5 shows the inliers of the detected planes from an indoor scan and the polygons computed by the α-shape for the plane with the most inliers (corresponding to the ceiling).

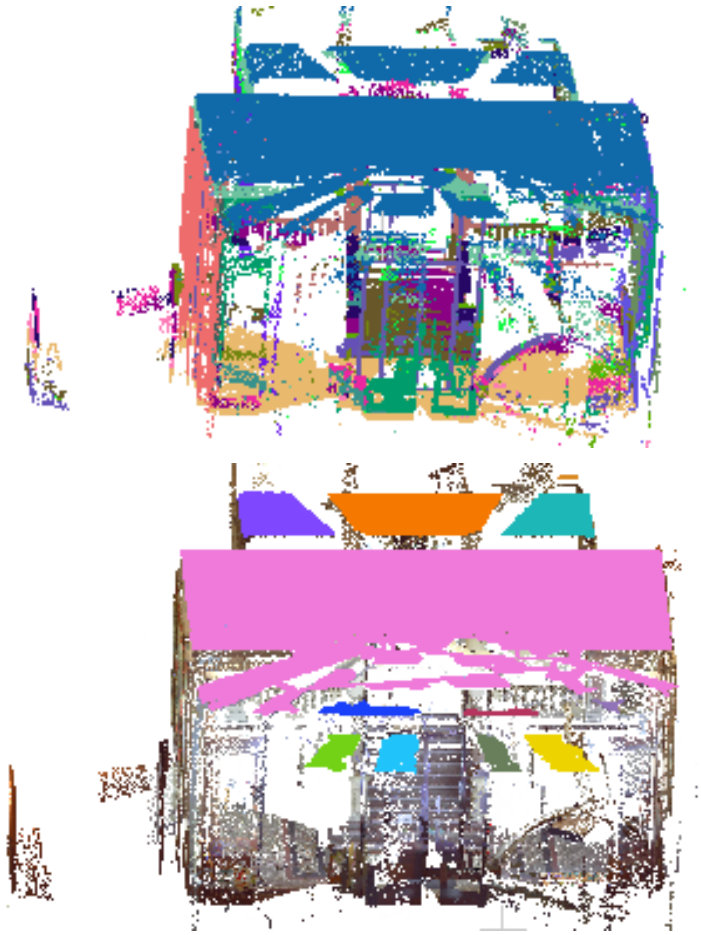


Figure 3.5: Top: Inliers of the estimated planes from an indoor scan, Bottom: Polygons extracted from the prominent plane.

3.4.3 Delaunay triangulation and Alpha shape

Delaunay triangulation:

Delaunay triangulation is the dual form of the Voronoi diagram that allows us to perform a unique division of space according to nearest neighbors. Following (Zhou and Yan, 2014), given a set of points $S\{P_i|i = 1, 2, ..., n\}$, the Voronoi diagram is the set of cells V_i defined by: $V_i = \{P|d(P, P_i) \leq d(P, P_j), \forall j \neq i\}$. AS d represents the distance, we can consider V_i the locus of the points closer to P_i than any other points

in S. Then, the Delaunay triangulation can be computed as the dual of the Voronoi diagram 3.6

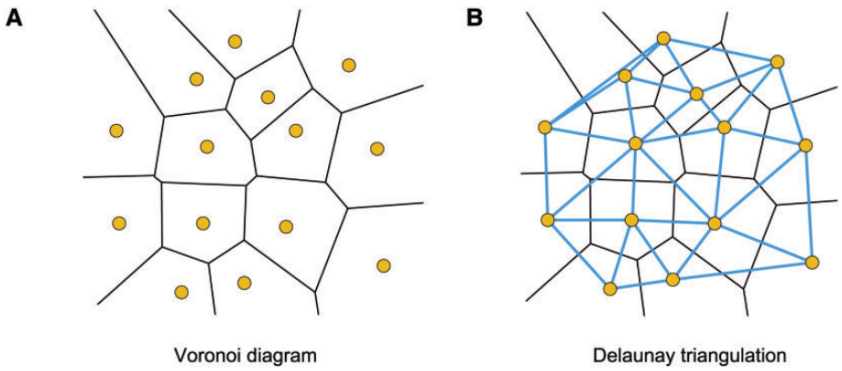


Figure 3.6: Construction of the Delaunay triangulation. (A) Voronoi diagram for a set of points. (B) Delaunay triangulation (Zhou and Yan, 2014).

Alpha shape

This technique has been proposed in (Edelsbrunner et al., 1983). It can be derived from Delaunay triangulation, which offers a concrete definition of a shape to represent the structure of a set of points. It represents a Delaunay triangulation (2D or 3D) for which each of the triangles or tetrahedron ("simplex") is compared to the value of an alpha (α) parameter determined by the Euclidean distance between the points 3.7.

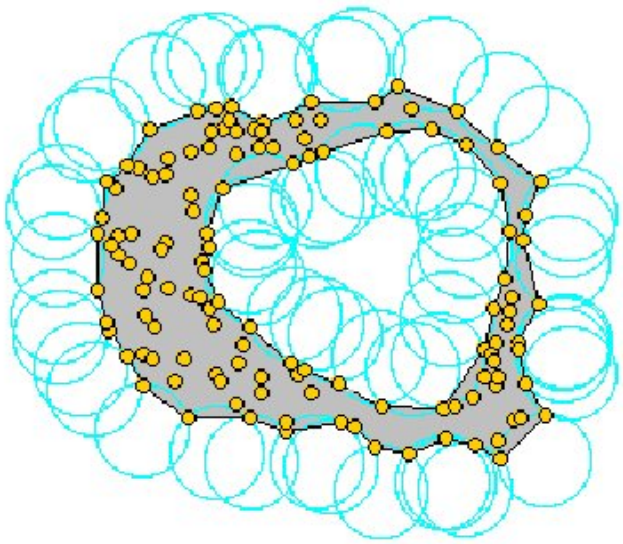


Figure 3.7: The basic idea of alpha shape (www.cgal.org)

3.5 3D line segment detection from LIDAR data

3.5.1 Introduction

The problem of detecting line segments has received much less attention in 3D point clouds compared to imagery where it has been well studied. With the current LiDAR devices, dense point clouds have become increasingly common, and the need for new approaches to manage them has grown more and more. Given the high density and the large volume of these point clouds, such processing requires a very large calculation time. So an abstraction step is necessary in order to reduce this complexity. Linear structures are considered among the best abstractions of LiDAR data especially in urban environment (Hofer et al., 2017).

3.5.2 State of the art

Several studies have focused on 3D line segment detection in 3D datasets. A fast and precise method to detect sharp edge features has been proposed in (Bazazian et al., 2015). This method works by analyzing the eigenvalues of the covariance matrix computed from each point's k-nearest neighbors. A new framework to extract line segments from large-scale point clouds has been proposed in (Lin et al., 2017). The key idea is to segment the input data into facets by performing local k-means clustering on carefully selected seeds. These facets provide sufficient information for determining linear features in the local planar region. The false positive detected line segments are filtered by introducing the concept of Number of False Alarms (NFA) into 3D point cloud context. The authors of (Hackel et al., 2016) presented a contour detector in unstructured 3D point clouds based on three main steps: First, a binary classifier is trained which, for each 3D point, predicts the likelihood of lying on a contour. Then, regularly spaced points with high contour scores are found and linked into a graph of candidate contours. Finally, an optimal subset of those candidates is selected as final wireframe edges, by approximate inference in a higher-order random field defined over the graph edges and their adjacency relations. An efficient algorithm has been proposed in (Lin et al., 2015), which is able to accurately extract line segments from large-scale unorganized raw scan point clouds.

The 3D line-support regions are also extracted at the same time and fitted by the Line-Segment-Half-Planes (LSHP) structure, which provides a geometric constraint for line segments, making their detection more reliable and accurate.

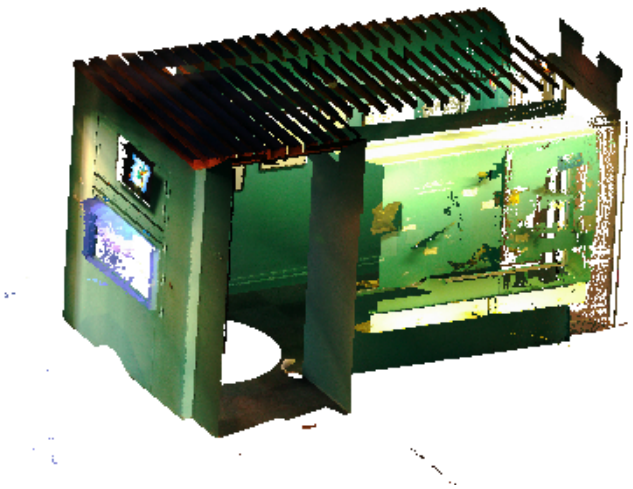
3.5.3 Classification of 3D line detection methods

Following (Lu et al., 2019), the problem of 3D line segment detection can be classified into three categories:

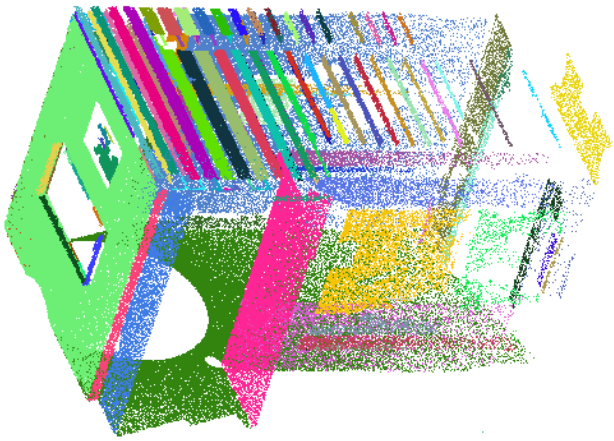
- Point based methods: use least square fitting of 3D line segments after detecting the boundary points. The main problem of this kind of method is the non-robustness to the noise.
- Plane based methods: a 3D line segment can be generated by the intersection of two 3D planes. The drawback is that the endpoints of the intersection line are generally difficult to determine.
- Image based methods start by converting the 3D point cloud into images. They then extract 2D line segments for each image. Finally, the 2D line segments are reprojected to the point cloud to get the final 3D line segments. The difficulty of this kind of method is that the adequate resolution of image rendering is hard to determine.

3.5.4 Selected method

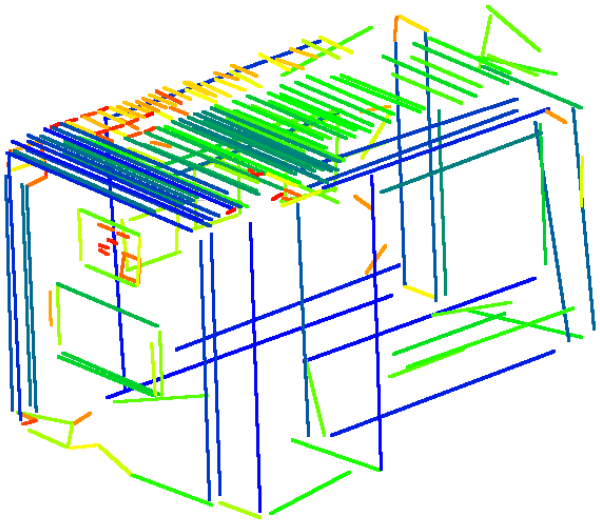
In this work, we have chosen the algorithm proposed in (Lu et al., 2019) to extract 3D line segments from the LiDAR data. It is a simple and efficient algorithm that starts by segmenting the point cloud into planar 3D regions via region growing and merging. All the points belonging to the same planar region are projected into the supporting plane of this region to form a 2D image. Then, 2D contour extraction and least square fitting are performed to detect 2D line segments. Finally, these 2D line segments are transformed back into the 3D frame to get the 3D segments as shown in figure 3.8. This algorithm belongs to the image based category.



(a)



(b)



(c)

Figure 3.8: 3D line segments detection from an indoor scan and colored according to the scale (where the scale of the point distribution in the neighbourhood of a given point p represents the distance between p and its third closest neighbouring point) : (a) Indoor point cloud, (b) Detected planes, (c) Extracted 3D lines.

3.6 3D lines segments detection and reconstruction from image data

3.6.1 Introduction

Due to the lack of texture in man-made environments such as indoor scenes, Structure from Motion (SfM) techniques based on detected salient points may fail or yield inaccurate results. To address these issues, a number of authors advocate the use of lines rather than points in such situations. Indeed, lines can generally be detected even in presence of uniform areas, at edges. Given the importance of lines to solve the problems of calibration, reconstruction and registration, several studies are oriented towards the proposal of efficient detectors for these geometric units. These studies range from the detection of 2D lines in a single image to the reconstruction of a 3D model from images sequence.

3.6.2 Line detection from an image

According to (Akinlar and Topal, 2011), traditional line segment detection algorithms can be summarized in the following steps:

- Computing an edge map, generally using Canny edge detector (Canny, 1986)
- Applying the Hough transform in order to extract all lines that contain a certain number of edge points.
- Breaking lines into segments using gap and length thresholds.

This traditional scheme is slow and leads to the generation of many false detections. To correct and overcome this problem, new families of algorithms based on *a contrario* theory has been proposed. These algorithms become currently state of art methods for line detection. Using *a contrario* theory, the algorithm can automatically define if a line is meaningful or not with a score called Number of False Alarms. The first proposed algorithms in this category are EDLines (Akinlar and Topal, 2011) and Line Segment Detector (LSD) (Von Gioi et al., 2012). After, a multi-scale extension of LSD (multiscale line segment detector (MLSD)) was proposed in (Salaün et al.,

2016).

EDLines: It is a real-time line segment detector. It is based on the edge segment chains produced by a novel edge detector, Edge Drawing (ED). This detector includes a line validation step applying the Helmholtz principle, which lets it control the number of false detections.

LSD: It is a linear-time Line Segment Detector which works on any digital image without parameter tuning. The idea is to detect segments as connected components of pixels with similar gradient direction. However, because of noise or lack of contrast, some components may be separated by a few pixels and the detector then breaks the line into several shorter segments. This causes an over-segmentation phenomenon, noticeable in high resolution images.

MLSD: It is a multi-scale extension to the popular LSD. Its multi-scale nature makes it much less prone to over-segmentation, more robust to low contrast and less sensitive to noise. Being based also on the *a contrario* theory, it retains the parameter-less advantage of LSD, at a moderate additional computation cost.

3.6.3 3D line segments reconstruction

3D reconstruction from images is a very active area of research. The first studies have been done through exploiting image feature-points and their invariant descriptors (e.g., SIFT (Lowe, 2004)). As in man-made environment, most of objects are structured and can be outlined by a bunch of line segments, a new generation of 3D line reconstruction methods is born. However, different from point-based reconstruction, reconstructed 3D lines may have large displacement from the ground truth pose in spite of a very small sum of reprojection errors. We distinguish two families of methods of 3D line segments reconstruction.

Pairwise reconstruction

For this family of methods, we only need two overlapping images to perform the reconstruction. According to (Baillard et al., 1999) and (Heuel and Forstner, 2001),

the two end points of a segment generate two epipolar lines in the other images. These two lines define a region, called the epipolar beam, which necessarily intersects or contains the corresponding segment. This reduces the complexity of the search for corresponding segments.

Multi-viewer reconstruction

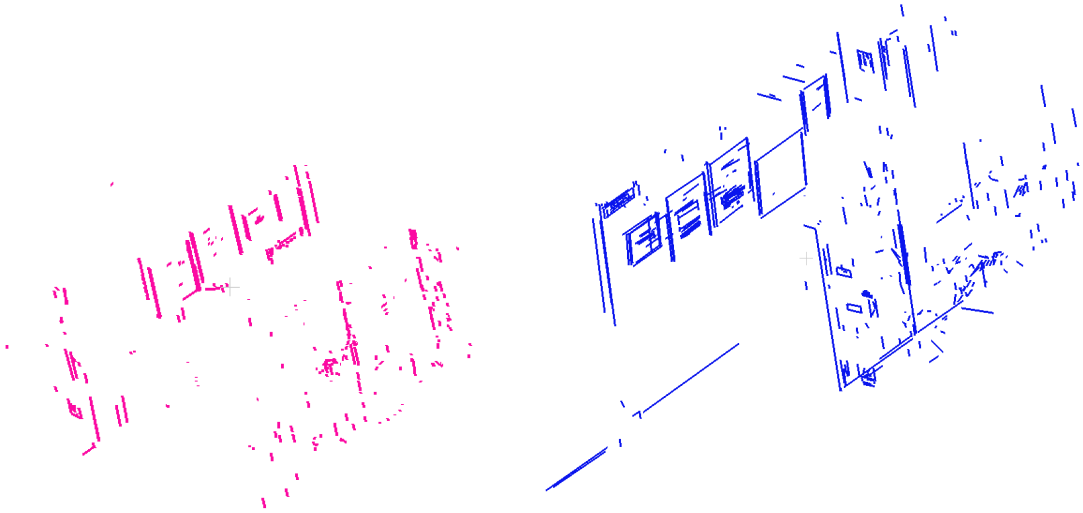
If more than two images are available, it is easy to extend the above pairwise reconstruction method to deal with multiple views reconstruction. A novel approach for the reconstruction of straight 3D line segments has been proposed in (Jain et al., 2010). The main idea is to impose global topological constraints given by connections between neighbouring lines to generate an improved reconstruction. Additionally, this approach does not employ explicit line matching between views, thus making it more robust against image noise and partial occlusion. It removes outliers by merging independent reconstructions, that are generated from different base images. A new method for reconstructing Line segments (LS) in structured scenes has been proposed in (Li et al., 2016). This method represents a potential solution to solve the uncertainties in 3D line segments reconstruction through LS match grouping, space plane estimation and image LS back-projection. By introducing a new minimal four-vector line representation based on Plücker coordinates, the authors of (Zhou et al., 2020) propose a method to analyze the uncertainty of line reconstruction. Each component of the compact representation has a certain physical meaning about the 3D line's orientation or position. The reliability of the reconstructed lines can be measured through the confidence interval of each component in the proposed representation. The article provides a quantitative evaluation of accuracy.

Selected method

We chose to work with the method developed in (Hofer et al., 2017). This method uses an oriented image sequence as input, whose camera poses can be obtained by any conventional Structure From Motion (SfM) pipeline. It establishes a large set of potential line correspondences between images using weak epipolar constraints and separates correct from incorrect matches for each segment individually using a scoring formulation based on mutual support. It generates a final line-based 3D model

by clustering 2D segments from different views, using an efficient graph-clustering formulation. We used the corresponding code available on `https://github.com/manhofer/Line3Dpp` to recover the 3D lines. This code uses LSD detector to detect 2D lines in images. To improve the performance of this algorithm we have replaced LSD by our more robust variant of MLSD. Figure 3.9 shows the impact of the applied modifications in the results.

(a)



(b) (c)

Figure 3.9: Reconstructed 3D line cloud: (a) An image from the 2D image sequence, (b) The reconstructed lines using LSD detected lines, (c) The reconstructed lines using MLSD detected lines.

3.7 Openings detection

3.7.1 Introduction

Recently, the topic of building reconstruction has drawn increasing interest. It has various applications, among them urban planning and virtual tourism. 3D city model represents a very important platform in a smart city. This platform is essential to produce environmental simulation and master planning applications. Detailed building facade structures can be recovered in high density point clouds acquired by Terrestrial Laser Scanning (TLS). The automated segmentation of building facades and the detection of their elements can reduce the effort of 3D reconstruction of existing buildings and even entire cities. Openings are considered the most important feature of facades and play a crucial role in generating photo-realistic building models. The similarity and repetitive patterns of these elements can help us to discover facade structures. These components are considered as premier cues for indoor outdoor registration. Therefore, they must be detected correctly and with great precision.

3.7.2 State of the art

Many works have dealt with the windows detection problem. A hole based windows extraction method was proposed in (Pu and Vosselman, 2007), where a segmentation process was performed in order to group laser points in planar segments. A TIN (Triangular Irregular Network) of the wall segments is generated after recognizing them. The windows points, extrusion points and door points are identified by classifying the hole points. An automatic approach to facade and window detection from mobile LiDAR Data by combining two strategies was proposed in (Wang et al., 2011). The first strategy used Principal Component Analysis (PCA) to cluster point clouds into potential facade regions, while the second extracts the facade by applying RANSAC (Random Sample Consensus) plane fitting to potential facade regions. Finally, the windows detection is achieved in two main steps: potential window point detection and window localization. The authors of (Tuttas and Stilla, 2011) have proposed a new method for detecting windows in airborne multi-aspect laser scanning data which represent a sparse point cloud. They started by detecting facade

planes using point normals and Region Growing (RG) algorithm. Then, the windows are detected using the indoor points (points lying behind the detected facade planes) which have a special interest due to their regularities in appearance. Another method for window and door detection has been proposed in (Nguatem et al., 2014), which applies a model selection based on Bayes factor and uses a Monte Carlo simulation to generate and score competing models. This method has proven its ability to distinguish between various window and door shapes. A robust algorithm to locate opening corners for terrestrial LiDAR points has been proposed in (Li et al., 2018). This algorithm searches and analyzes the facade with a globally wide sliding window that reduces the influence of noise in horizontal and vertical directions. The changes in facade elements were inspected by computing gradients in each sliding window, applying statistical analysis and deriving the symmetry information. The authors of (Recky and Leberl, 2010) have proposed a modified gradient projection method designed for automatic processing of complex building facades. This algorithm is able to detect many different windows types without a learning step. An automatic method to detect windows and estimate their parameters in 3D LiDAR point clouds has been proposed in (Aijazi et al., 2014). This method fuses information from both symmetrical and temporal correspondences to estimate window features. A new algorithm for openings detection in indoor point clouds has been proposed in (Assi, Landes, Macher and Grussenmeyer, 2019). Holes corresponding to openings were detected after identifying walls in 2D binary image. Each opening was submitted to an energy function with two terms: data and coherence. Those functions help the authors to determine if an opening is due to a window/door or an object obstructing the acquisition.

3.7.3 Proposed method

According to (Pu and Vosselman, 2007), we can define the following characteristics for an opening:

- Most openings have a rectangular vertical shape of limited extent (a few meters) positioned within a vertical plane (the wall).
- Openings which are not covered with curtains cause holes on the wall segments

where the Laser beams traverse the wall plane.

- Openings which are covered with curtains reflect the Laser beam which produces LiDAR points which are not on the wall plane.

According to these properties we propose a robust method to detect openings in LiDAR data compatible with any acquisition mode (mobile outdoor or static indoor) which is summarized in figure 3.10.

LiDAR data segmentation and facade selection

RANSAC is the most used method for LiDAR data segmentation. But, it has some limitations such as under-segmentation (grouping points from nearly co-planar but disjoint surfaces in a single planar primitive). In order to overcome the limitation of RANSAC, we chose to use the two solutions proposed in Section 3.4.2 depending on the nature of the data and the acquisition platform.

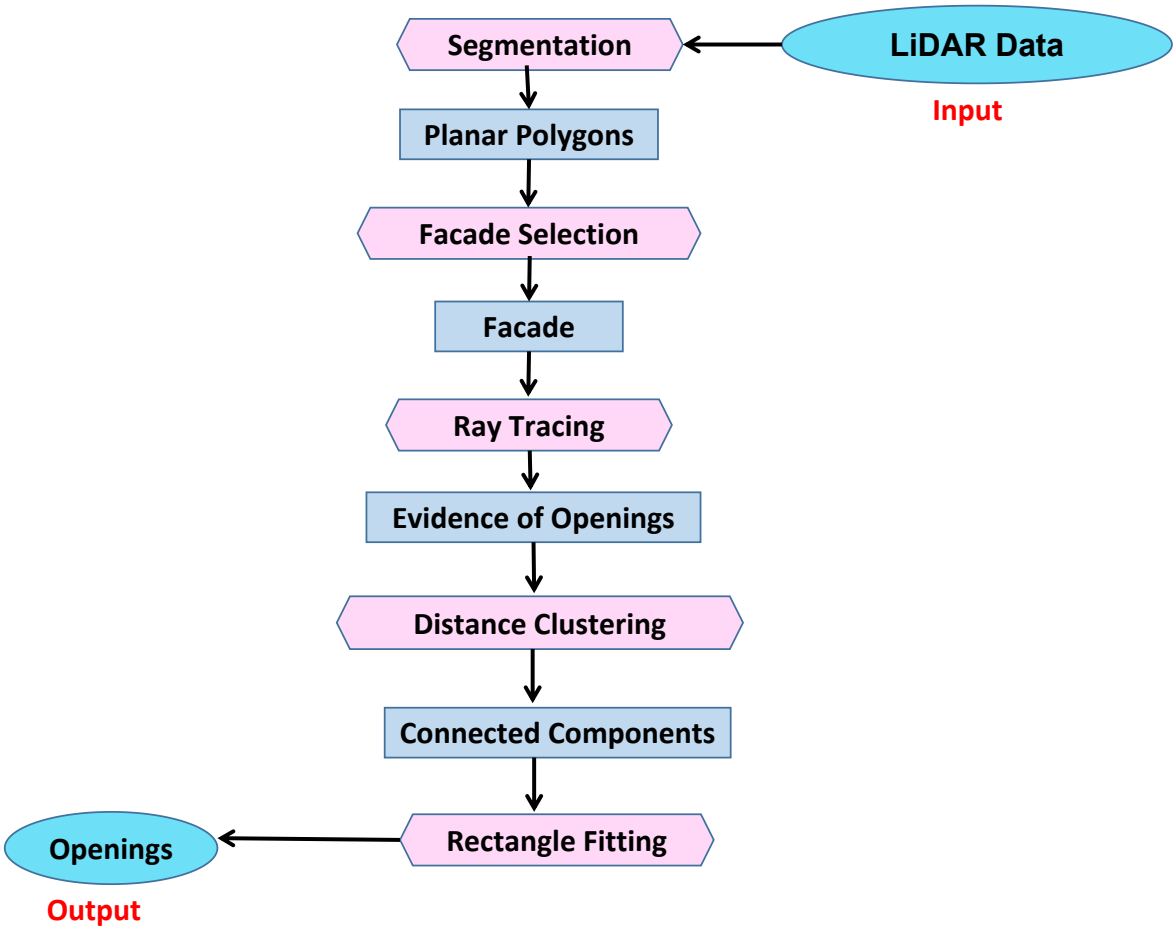


Figure 3.10: Openings detection pipeline

After the planar polygons extraction, we have selected the facades as large vertical

polygons.

Evidence of openings detection

As the LiDAR beams usually cross the facade through openings, we propose to start by detecting the evidences of openings as the intersection points of these beams and the detected facades using Ray Tracing:

1. For each point P_i, we trace a ray R_i from the LiDAR optical center O to P_i.
2. We find the intersection points P_i^j of R_i with the supporting planes $\mathcal{P}_j$ of each façade polygon F_j.
3. If one P_i^j lies inside the polygon and the distance between P_i and $\mathcal{P}_j > d_{min}$, we add P_i^j to the list E_j of evidences of openings on wall j. In Section 3.7.3, we use $d_{min} = 0.1\,\text{m}$ in order to exclude noisy points that may still represent points of the facade.

Outline openings extraction

For each wall plane $\mathcal{P}_j$ the evidences of openings E_j are grouped in vertical rectangles:

- Extract the connected components of E_j with a distance threshold.
- Estimate the minimum bounding rectangle of each component.
- For each rectangle, transform each 2D corner to a 3D point and create the four 3D line segments corresponding to its edges to get a 3D representation of our shape as shown in Figure 3.11.

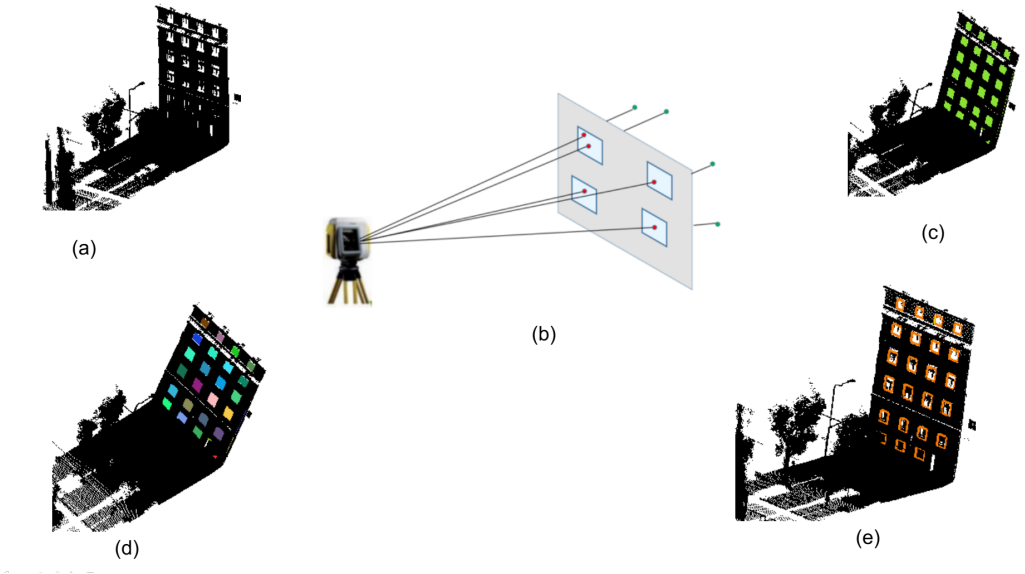


Figure 3.11: Illustration of opening detection steps on an outdoor scan: (a) initial outdoor scan. (b) principle of opening evidence detection. (c) detected openings evidence (green points). (d) connected components of opening evidence. (e) openings outlines (in orange).

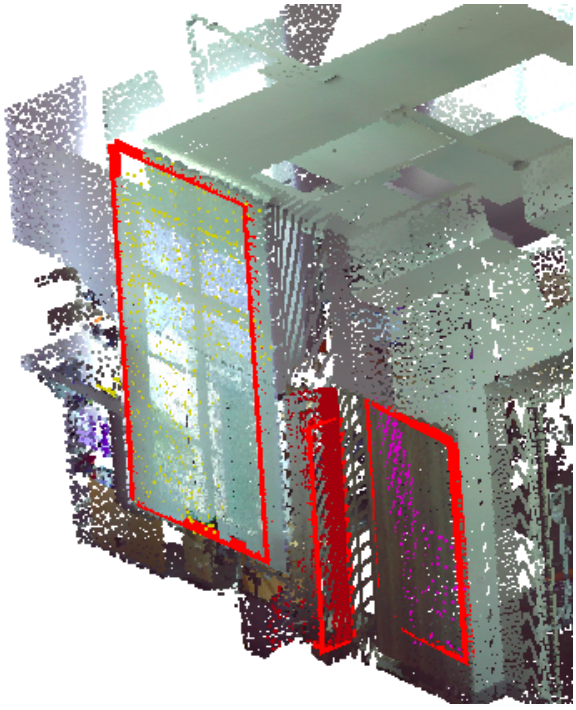


Figure 3.12: Openings detection in an indoor scan.

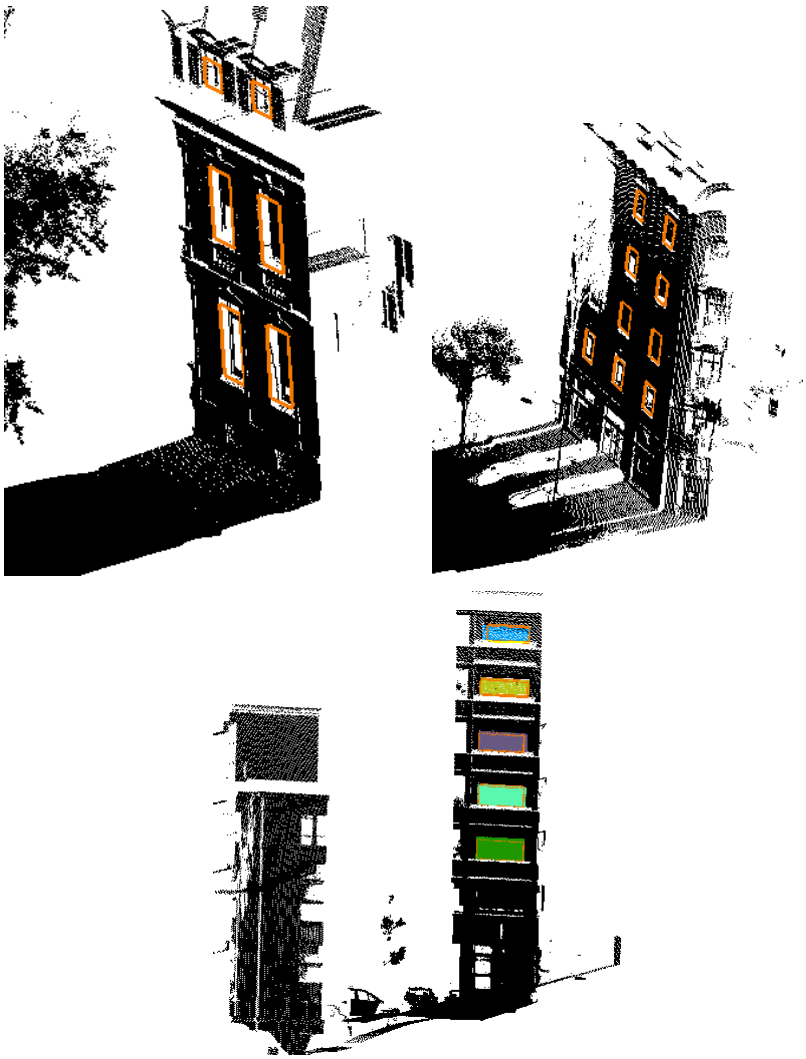


Figure 3.13: Openings detection results in several outdoor scans.

3.8 Conclusion

In this chapter we have tackled the problem of detecting geometric primitives which represent the core of data processing. The building environment has the following properties

- Most objects are composed of planar surfaces delimited by straight lines.
- We can present a building as a set of facades, where each one is composed of a set of openings.

Given properties above, our study was focused on the classification of primitives extraction methods followed by the following parts:

- Planar polygon extraction from LiDAR point clouds.

- 3D Line segments detection from LiDAR point clouds.
- Line segments detection and reconstruction from image sequences.
- Openings detection from LiDAR point clouds.

For each part, we started with a bibliographic study of existing works followed by the description of our proposed or selected methods. The obtained results show the efficiency of our methods (proposed or selected) to extract the different geometric primitives. These geometric primitives will be used as basic elements for more complex processes. These processes will be studied in detail in the next chapters.

Chapter 4

Indoor/Outdoor Registration

4.1 Introduction

The registration of indoor and outdoor scans with a precision reaching the level of geometric noise represents a major challenge for Indoor/Outdoor building modeling. The lack of overlap between indoor and outdoor data is the most encountered obstacle, more so when both data sets are acquired separately and using different types of sensors. State of the art approaches generally deal with either the indoor or the outdoor, and often use strong priors of parallelism and orthogonality that are not necessarily verified. To achieve the desired accuracy, clues must be extracted from internal and external data and then matched. It is very restrictive because these clues should be visible from the inside and the outside. In this work, we explored fundamental properties of our data to propose potential solutions for this challenging problem. Given the specificity of the Lidar data that crosses the windows, we have performed an extraction of the planar polygons that offer us a compact and localized representation through the detection of the facades (vertical planar polygons with sufficient extent). Then, we used them to propose an efficient registration algorithm based on planar polygons as the first solution.

As an opening in the façade is the unique common entity that can be seen from inside and outside, it can help the registration of indoor and outdoor point clouds. So, we have integrated openings in a registration framework in order to introduce a second

solution for the indoor/outdoor registration problem.

The two proposed solutions (planar polygons based algorithm and openings based algorithm) are very effective for realizing the indoor/outdoor registration. But they have some limitations. in order to overcome these limitations, we have proposed a hybrid solution, which represents a combination of both. In this work, we assume the outdoor (mobile mapping) scan to be the reference as its georeferencing system has direct access to GNSS data, so our problem is to register the indoor scans together and with the outdoor scan.

4.2 State of the art

4.2.1 3D Point clouds registration

Integrating multi-platform, multi-angle, and multi-temporal data is a core step in several fields of research such as computer vision, remote sensing and 3D modeling. To achieve this integration, a registration step is necessary. For Terrestrial Laser Scanning (TLS), there are two main families of registration methods depending on the number N of input point clouds: pairwise ($N = 2$) and multi-view ($N > 2$) registration. In this work we are interested in the pairwise category. The most used approach to solve this problem is feature-based registration which can be decomposed in three main steps: extracting geometric features (points, curves, planes, and/or surfaces), selecting the correspondences, and using selected correspondences to estimate the optimal transformation (Dong et al., 2018). The Iterative Closest Point (ICP) algorithm is considered as the most used approach in the registration of point clouds (Besl, 1992). ICP starts with two overlapping point clouds and initial guess. The transformation parameters can be iteratively estimated by generating pairs of corresponding points and minimizing the error metric. The major disadvantage of this method is the convergence towards a local solution if the initial data are not spatially close or if the initial transformation is poorly estimated.

Several variants of the ICP algorithm have been proposed to improve its robustness such as using a point-to-plane error metric. At each iteration of the algorithm, the relative pose that gives the minimal point-to-plane error is usually estimated using

a standard nonlinear least-squares method. (Low, 2004) proposes an approximation of the nonlinear optimization problem with a linear least-squares one that can be solved more efficiently. An extension of the ICP framework to nonrigid registration that uses the same convergence properties of the original algorithm was proposed in (Amberg et al., 2007). Another approach (Sharp et al., 2002) proposes to use Euclidean invariant features in a generalization of ICP registration of range images. To find the correspondence of 3D range camera, the authors proposed to use either spherical harmonics or the second order momentum. Another method for detecting uncertainty in pose has been introduced in (Gelfand et al., 2003), where the transformations that can cause unstable sliding in the ICP algorithm have been estimated using a sampling strategy and the points that best contain this sliding have been picked. In order to minimize the search space for correspondence between two point clouds and to increase the accuracy of the registration, the authors of (Rabbani et al., 2007) have used standard primitives, namely planar and cylindrical surfaces, to introduce two solutions suitable in industrial environments: an indirect solution which separates object fitting and registration steps, a direct solution which simultaneously determines the shape and pose parameters of the objects as well as the registration parameters.

A Normal Distribution Transform (NDT) algorithm and its variants are proposed to deal with the problem of 3D points cloud registration. A certain cost is obtained that evaluates the alignment quality between the two point clouds, according to the desired transformation between the two point clouds and using a set of Gaussian distributions with different probability density functions (PDFs) to represent the referenced point clouds (Dong et al., 2020) (Magnusson et al., 2007). A registration workflow which simulates the well-known RANSAC was proposed in (Al-Durgham et al., 2013) to register overlapping terrestrial laser scans by using 3D linear features to establish an automatic matching strategy.

The authors of (Theiler et al., 2012) have used virtual tie points generated by the intersection of three non-parallel planes in two different scans. So, the objective of the registration algorithm is to search for the assignment which preserves the geometric configuration of the largest possible subset of all tie points.

The authors of (Forstner and Khoshelham, 2017) have proposed methods to efficiently register point clouds by introducing new optimal and sub optimal direct solutions based on plane-to-plane correspondences for determining the relative motion. A fast 3D point cloud registration in cluttered urban environments has been proposed in (Xiao et al., 2012), where the registration is defined as a correlation problem. Other works based on deep learning techniques have been proposed in the literature, such as the method proposed in (Elbaz et al., 2017), where the authors have introduced an innovative point cloud registration algorithm based on two original approaches. The first consists in using super-points (selected by a random sphere cover set) as the basic units for matching, instead of key-points. Whereas the second consists in encoding local 3D geometric structures using a deep neural network auto-encoder.

4.2.2 Indoor/Outdoor registration

The registration of indoor and outdoor scans is a challenging problem for Indoor/Outdoor building modeling. Two registration and georeferencing methods for exterior and interior point clouds, in the case of heritage building recording, are tested in (Murtiyoso and Grussenmeyer, 2018). The first corresponds to the independent georeferencing method, in which the indoor and outdoor scans were georeferenced separately. The second is a keypoint based registration method. These tie points concern mainly openings such as windows, which are evenly distributed throughout the building. Windows are manually segmented in the indoor and outdoor point clouds. Then keypoints are detected for both sides of the windows, resulting in two sets of keypoints. The work presented in (Assi, Landes, Murtiyoso and Grussenmeyer, 2019) has proven that the use of key points rather than the entirety of a point cloud for an ICP algorithm can improve its performance to register indoor and outdoor scans. An approach for automatically aligning the non-overlapping indoor and outdoor parts of a 3D building model computed from image based 3D reconstructions has been proposed in (Koch et al., 2016). This alignment has been performed by identifying corresponding 3D structures that are part of the interior and exterior model (e.g. openings like windows).

4.3 Data

The data used for this study was acquired by two different means (mobile LiDAR scan for the outside and static scans for the inside) on the Zoological Museum of Strasbourg.

4.3.1 Outdoor data

The outdoor data used to experiment our method is a Mobile Laser Scanning (MLS) (cf Fig. 4.1 acquired with the Stéréopolis II Mobile Mapping System (MMS) (Paparoditis et al., 2012). The acquisition system gives access to the sensor topology inherent to such MLS acquisitions, that is usually lost during export to formats such as .las or .ply. The data was collected from three streets, north, south and east of the museum, the west façade facing a park inaccessible to the MMS. Each outdoor scan contains approximately 3 million points.

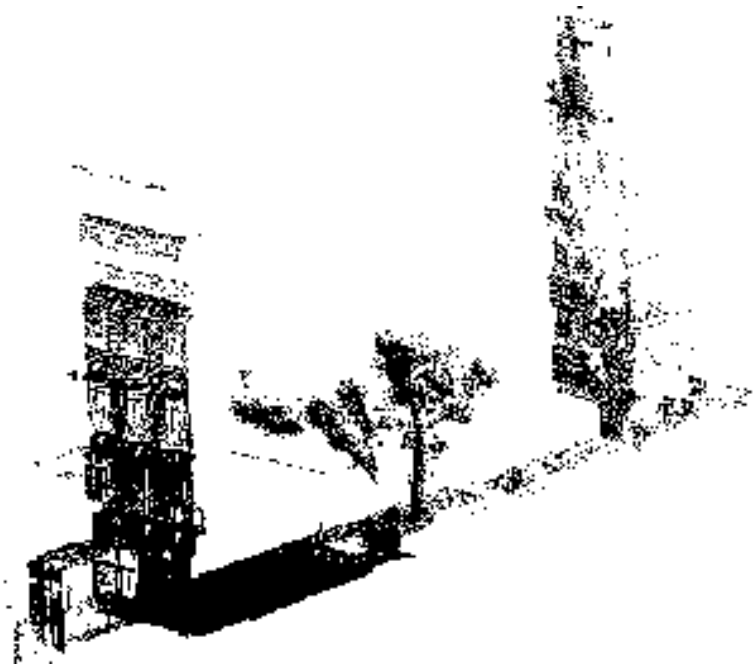


Figure 4.1: Outdoor MLS scan acquired with an MMS

4.3.2 Indoor data

The indoor data used in our study is composed of 30 static LiDAR scans of the inside of the Musée Zoologique, one or two per room. Each scan consists of roughly 500 million points, which were downsampled to around 2 million points for practical

reasons as shown in Figures 4.2 and 4.3. The Figure 4.4 shows some indoor and outdoor scans acquired in static mode at the zoological museum of Strasbourg. The indoor acquisition was performed using the *FARO Laser scanner Focus3D X 330*.

FARO Laser scanner Focus3D X 330: According to (Kallisto, 2022), *FARO Laser scanner Focus3D X 330* is ideal for both indoor and outdoor surveying applications. It offers everything you would expect from a professional laser scanner. The *FARO Laser scanner Focus3D X 330* has an extra long range of 330 m. It is equipped with a GPS and adapted for outdoor scanning, even in strong sunlight. Its accuracy is of the order of $\pm$ 2 mm and its acquisition speed reaches up to 976,000 points/second.

Figure 4.2: Indoor scan acquired in static mode inside the Zoological Museum of Strasbourg.

Figure 4.3: Indoor scans of the ground floor of the Zoological Museum of Strasbourg.

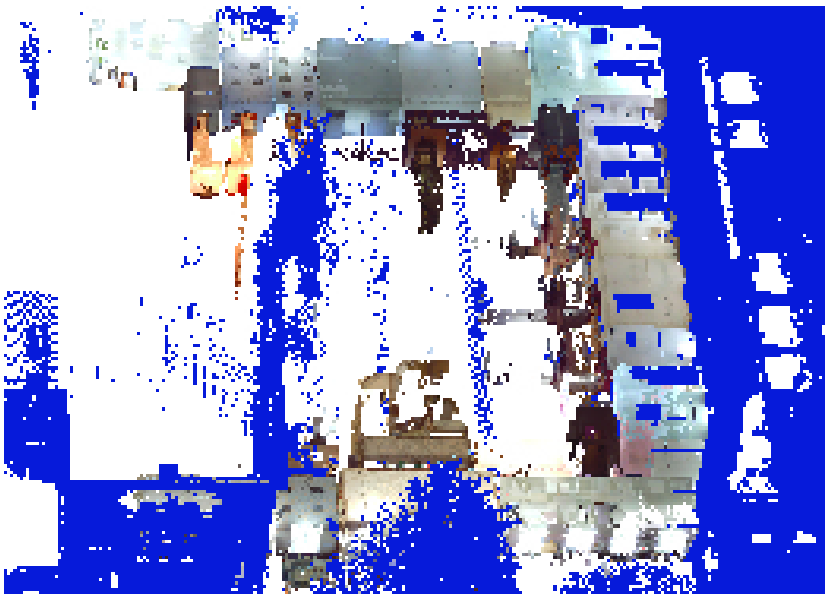


Figure 4.4: Indoor and outdoor scans acquired in static mode at the Zoological Museum of Strasbourg.

4.4 Planar polygons based registration

The work carried out has confirmed that the environment and the type of data drive the choice of the registration algorithm. For example, in man-made environments, where most objects are bounded by planar surfaces, the ideal is to choose a method of registration based on plane correspondences (Theiler et al., 2012) or primitive correspondences. The basic idea of our contribution is to perform an extraction of the planar polygons, from both the indoor and outdoor data, then to match them by clustering polygons according to their normal direction, then by their offset in the normal direction. We use this clustering to find possible polygon correspondences

(hypotheses) and estimate the optimal transformation for each hypothesis. Finally, a global robust energy is computed for each hypothesis in order to select the best one as shown in Figure 4.5. The polygon extraction step has been carried out using the methods proposed in Section 3.4.2. The interest of polygons relative to planes is that they have a spatial extent limited to the areas where they have supporting points in the input data, so they form a simple and compact summary of our LiDAR scans. Given the specificity of LiDAR data that pass through windows, we propose to start with the extraction of the buildings' interior points captured from external scans; these are the points where the laser ray crosses the façades through apertures, mostly windows. Afterwards, we perform the registration.

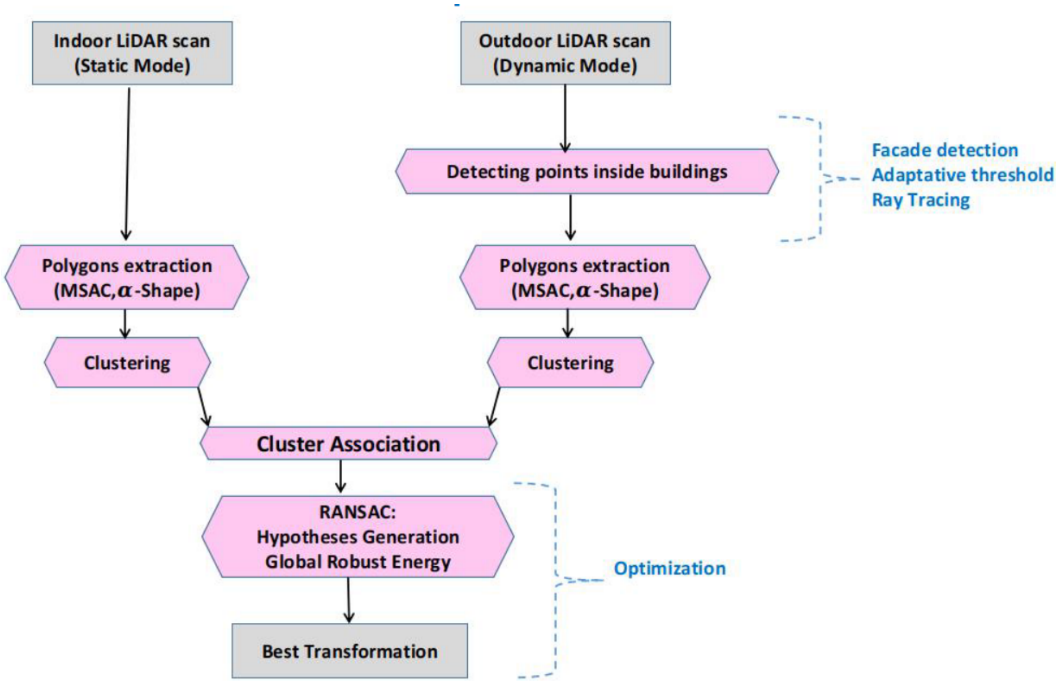


Figure 4.5: Pipeline Details of planar polygons based registration algorithm.

4.4.1 Detecting points inside buildings in the outside scan

As the LiDAR beam usually passes through windows as shown in Figure 4.6, we propose to start by detecting façades from the outdoor scan, then detecting indoor points as points behind a façade by ray tracing.

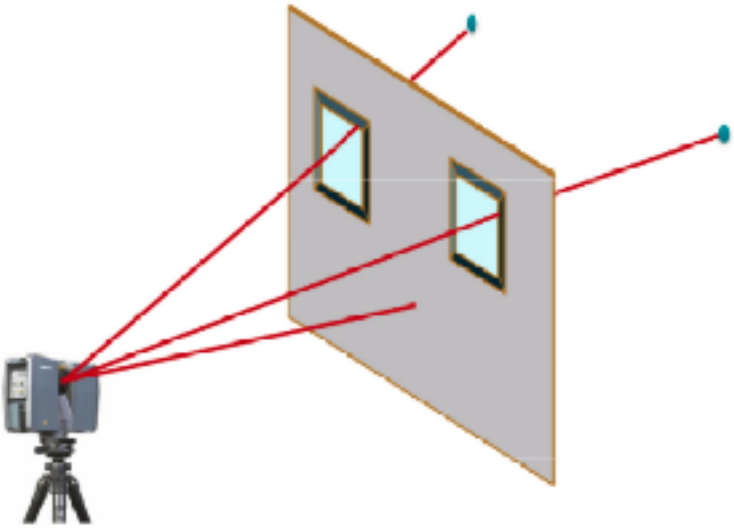


Figure 4.6: LiDAR rays that pass through windows.

Façade detection

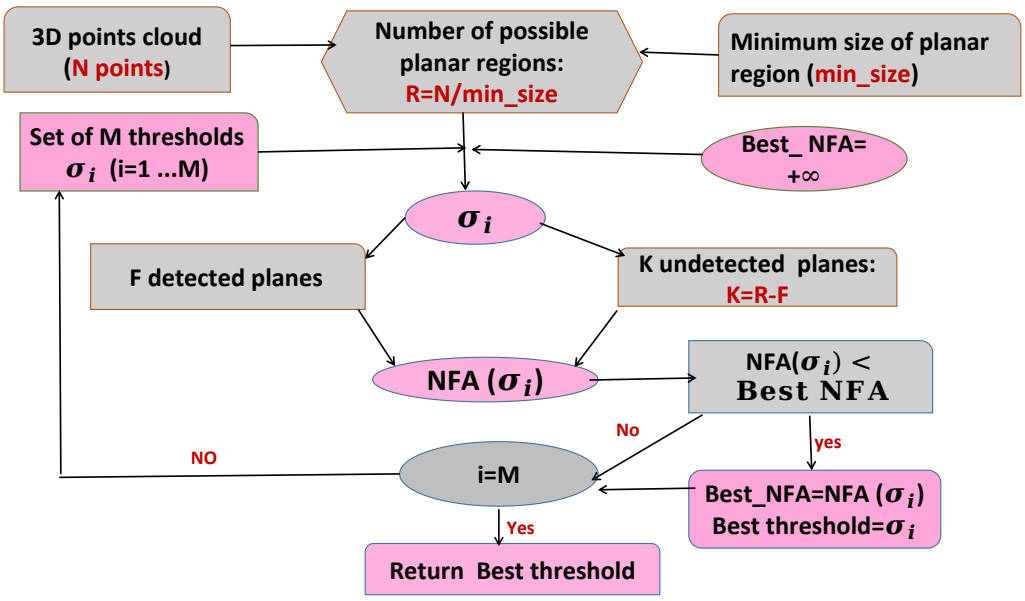


Figure 4.7: Design of the *contrario validation method*.

For façade detection, we decided to use the planar polygon detection of Section 3.4.2 (RANSAC based on sensor topology) to detect the façades as large vertical polygons. Thus we will keep only the detected planar polygons which are sufficiently vertical (deviation below $3°$).

The choice of the threshold on inlier distance is a crucial factor as a bad choice can cause important under or over detection of façades, leading to a bad detection of interior points. In order to select automatically an appropriate threshold, we propose an *a contrario* algorithm that takes as input M possible thresholds and selects the

threshold with the lowest NFA (Number of False Alarms). Given a 3D point cloud of N points, the definition of the minimum size $\min_{\text{size}}$ of a planar region allows us to deduce the maximum number of possible planar regions:

$$R = \frac{N}{\min_{\text{size}}} \tag{4.1}$$

For each given threshold, RANSAC estimates $F \leq R$ planes, so at most $K = R - F$ planes are undetected as shown in Figure 4.7. Inspired by (Bughin and Almansa, 2010), we can define the error as being the probability that a significant existing planar region is not detected. Another possibility is the probability that the number of detected planes is less than the number of undetected planes:

$$NFA(\sigma) = N_{\text{iter}} \times p(K \geq F \mid \sigma) \tag{4.2}$$

The probability $p(K \geq F \mid \sigma)$, can then be upper bounded by the tail of the binomial law of parameter P defined by:

$$P(\sigma) = \frac{2\sigma \times diag^2}{v} \tag{4.3}$$

where $diag$ is the diagonal diameter of the bounding box of the 3D point cloud and V its volume. Then:

$$p(K \geq F \mid \sigma) \leq \sum_{j=F}^{K} \binom{K}{j} P(\sigma)^j \left(1 - P(\sigma)\right)^{K-j} \tag{4.4}$$

Ray Tracing

Once the façades are detected, we detect indoor points by ray tracing:

1. For each point P_i, trace a ray R_i from the LiDAR optical center to P_i.
2. Find the intersection points P_i^j of R_i with the supporting planes $\mathcal{P}_j$ of each façade polygon F_j.
3. Test if P_i^j is inside the polygon (using the CGAL library (`https://www.cgal.org`)).

4. In order to remove ouliers close to the façade plane, we additionaly require the point to be at least 1 m away from the intersected polygon: ($dist(P_i, P_i^j) > 1m$).

5. If one P_i^j satisfies both criteria (inside the polygon and sufficiently far from the façade), tag it as an indoor point.

A result of this indoor point detection method is presented in Figure 4.8. Finally, the polygon detection algorithm of Section 3.4.2 (polygon MSAC) is run only on the points detected as indoor from the outdoor scan, yielding a limited number of polygons.

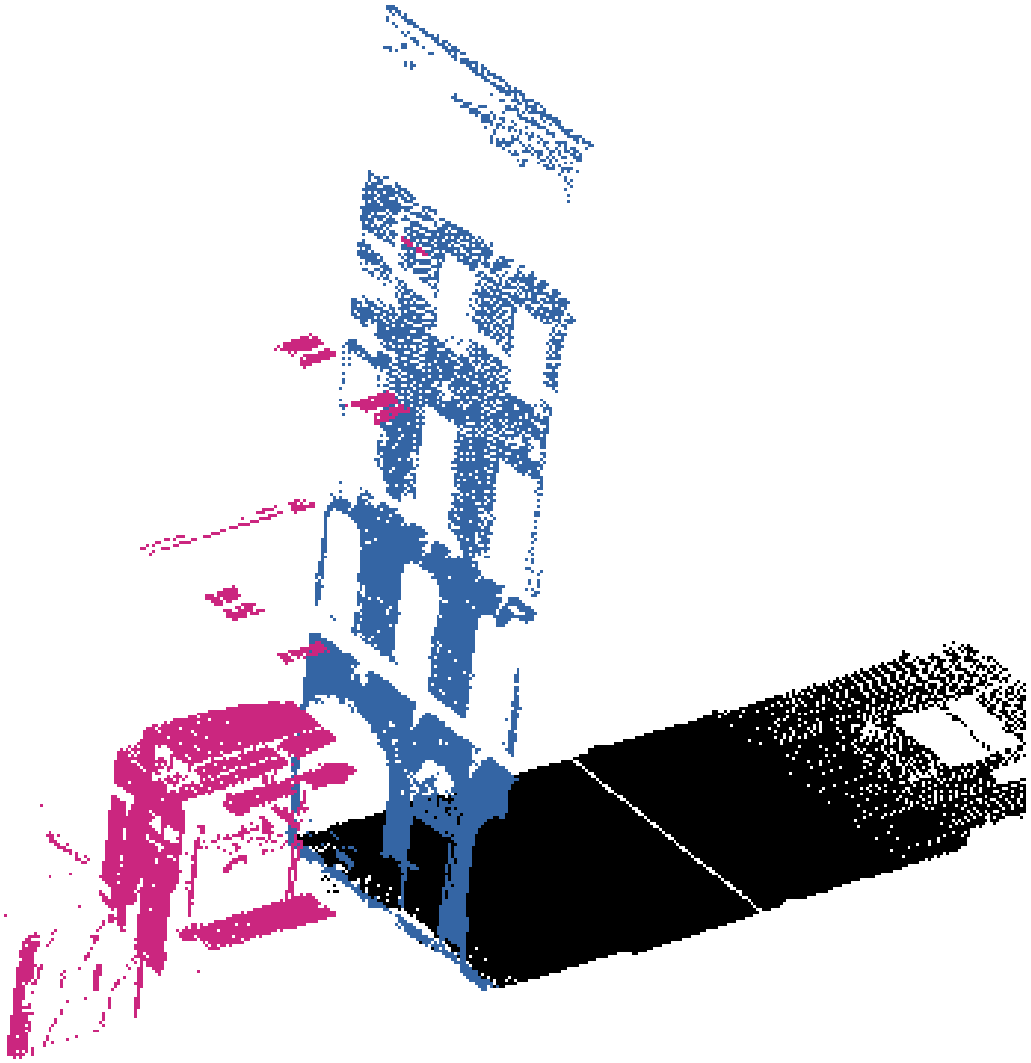


Figure 4.8: Indoor points detected from an outdoor scan, pink: the detected points, blue : inliers of the vertical plane that represents the façade

4.4.2 Matching planar polygons

Selection of correspondences is a crucial step for the registration. If we have at least three correct correspondences of polygons with independent normals, it is possible to find the relative rotation/translation between the indoor and outdoor scans. We will start by presenting a simple threshold based matching in order to introduce the three main criteria used to match polygons, then propose a more robust matching based on clustering algorithm.

Threshold based matching

Let us call:

- P_1 and P_2 detected planar 3D polygons from the two LiDAR scans: indoor scan $(Scan_1)$ and outdoor scan$(Scan_2)$.
- $\mathbf{n}_1$ and $\mathbf{n}_2$ the corresponding plane normals.
- $\mathbf{g}_1$ and $\mathbf{g}_2$ the corresponding centroids.
- $\mathcal{P}_{1,2}$ their bisector plane.
- P_1' and P_2' the projections of P_1 and P_2 on the plane $\mathcal{P}_{1,2}$.

We propose first a simple filter based on three measures:

1. Angle: $Angle(P_1, P_2) = \langle \mathbf{n}_1, \mathbf{n}_2 \rangle$
2. Distance: as there is not a standard definition for this distance, we have chosen to define it as the sum of distances from the centroid of each polygon to the bisector plane:

$$Distance(P_1, P_2) = dist(\mathbf{g}_1, \mathcal{P}_{1,2}) + dist(\mathbf{g}_2, \mathcal{P}_{1,2}) \tag{4.5}$$

3. Overlap:

$$Overlap(P_1, P_2) = \frac{|P_1' \cap P_2'|}{\min\left(|P_1'|, |P_2'|\right)} \tag{4.6}$$

where $|.|$ denotes the area. Note that we did not use the union on the denominator because a planar part of the inside scene is seen through an opening from

the outside so only a very limited portion of it can be detected which would result in very low overlap with a more standard definition.

In practice, finding appropriate thresholds for these three criteria is tedious and leads often to multiple or no matchings. This is why we propose a more robust approach.

Cluster based matching

We propose a more robust matching for polygons based on three main principles: clustering the polygons (by direction then by offset), enumerating match hypotheses between the clusters and evaluating which hypothesis is the best.

Algorithm 1 Greedy direction clustering

Input: P_i a set of planes with normals $\mathbf{n}_i$ and number of inliers n_i. ϵ a tolerance angle (typically $\pi/4$ rad).
Clusters initialization:

- $C_1 = \{P_1\}$ where P_1 is the plane with most inliers
- $C_2 = \{P_2\}$ where P_2 is the plane with most inliers among planes for which $\mathbf{n}_i.\mathbf{n}_1 < cos(\epsilon)$
- $C_3 = \{P_3\}$ where P_3 is the plane with most inliers among planes for which $\mathbf{n}_i.\mathbf{n}_1 < cos(\epsilon)$ and $\mathbf{n}_i.\mathbf{n}_2 < cos(\epsilon)$

Mark P_1, P_2 and P_3 as processed and all other P_i as unprocessed
Each cluster C_k has a normal $\vec{c}_k$ computed as the weighted centroid of all the normals of the planes in C_k
Let P_{cur} be the unprocessed plane with the most inliers. Mark P_{cur} as processed.
compute $k_{min} = \arg\min_k 1 - |\vec{n}_i.\vec{c}_k|$
compute $d_i = 1 - |\vec{n}_i.\vec{c}_{k_{min}}|$
If $d_i < \epsilon$, add P_i to the cluster $C_{k_{min}}$.
Call C^h and tag as horizontal the cluster for which $1 - |\vec{z}.\vec{c}_k|$ is minimum.
Call C^{v_1} and C^{v_2} and tag as vertical the two remaining clusters.

Direction clustering: For each input scan, we greedily cluster planes P_i according to their normals $\vec{n}_i$ by decreasing number of inliers n_i as detailed in Algorithm 1 that produces 3 clusters C^h, C^{v_1} and C^{v_2} for each scan.

Direction cluster matching We associate C^h for the two scans. We associate each remaining cluster (C^{v_1} and C^{v_1}) to the vertical cluster from the other scan with the smallest angle.

Alignment: we compute the rotation that aligns the three clusters: the vertical C^{v_1} and C^{v_2} and the horizontal clusters using the method described in Section 4.4.3

Hypotheses enumeration: for each associated direction cluster, we enumerate possible plane matches:

- For each of the three associated main cluster pair $A_m = (C^1_{k_1}, C^2_{k_2})$ we will call I_m the set of pairs of planes $(P^1_{i_1} \in C^1_{k_1}, P^2_{i_2} \in C^2_{k_2})$
- for each (J_1, J_2, J_3) in $I_1 \times I_2 \times I_3$, compute the translation that aligns the planes in (J_1, J_2, J_3) using the method described in Section 4.4.4.
- note that I_3 is extracted from the horizontal clusters which cover the displacement along z, I_2 is extracted from the vertical clusters which have the smallest angle with the y axis and I_1 is extracted from the vertical clusters which have the smallest angle with the X axis.
- keep the translation calculated by the best hypothesis.

Selection of the best hypothesis: To asses which hypothesis is the best, we define an energy that (1) is robust to outliers as many planes detected in one scan have no counterpart in the other (2) favors important polygon overlaps and (3) favors small distances over these overlapping parts. For robustness, we need a distance threshold d_{thr} above which a polygon is just considered as an outlier. Then for two polygons P_1, P_2 with centroids O_1, O_2 in the same cluster, we define a robust error:

$$E^{d_{thr}}(P_1, Scan_2) =$$

$$\frac{1}{d^2_{thr}} \int_{P \in P_1} min(d^2_{thr}, dist(P, Scan_2)^2) dP \tag{4.7}$$

We see that the energy is 0 when P_1 is completely overlapped by polygons of $Scan_2$ with a distance of 0 (the perfect case) and increases as distance augments and overlap decreases up to $|P_1|$ when the overlap is empty and/or the distance is over d_{thr}. In practice we make the approximation:

$$E^{d_{thr}}(P_1, Scan_2) \approx$$

$$\sum_{P_2 \in Scan_2} |P^2_1 \cap P^1_2| \frac{max(0, d^2_{thr} - Distance(P_1, P_2)^2)}{d^2_{thr}} \tag{4.8}$$

Note that we do not use our modified relative $Overlap$ function anymore as a large

overlap surface should be favored to a small one as it correponds to more input points. Finally, we can write our global robust energy as:

$$E_R(Scan_1, Scan2) = \sum_{P_1 \in Scan_1} E^{d_{thr}}(P_1, Scan_2) \tag{4.9}$$

We consider that the best hypothesis is the one that maximizes this energy.

4.4.3 Rotation estimation

Assuming that we have an association between three directional clusters of two scans:

- C_1^h and C_2^h the horizontal clusters of scans 1 and 2, $\vec{c_1^h}$ and $\vec{c_2^h}$ their normals.
- $C_1^{v_1}$ the first vertical cluster of scan 1 and $C_2^{v_1}$ the associated vertical cluster of scan 2, $\vec{c_1^{v_1}}$ and $\vec{c_2^{v_1}}$ their normals.
- $C_1^{v_2}$ the second vertical cluster of scan 1 and $C_2^{v_2}$ the associated vertical cluster of scan 2, $\vec{c_1^{v_2}}$ and $\vec{c_2^{v_2}}$ their normals.

We want to find the rotation that best aligns these three pairs of directions. Let $\vec{u}_1 = \vec{c_1^h}$ and $\vec{u}_2 = \vec{c_2^h}$. We are looking for the two vectors:

$$\{\vec{v}_1, \vec{v}_2\} = \underset{\vec{v}_1 \in \{\vec{c_1^{v_1}}, \vec{c_1^{v_2}}\}, \vec{v}_2 \in \{\vec{c_2^{v_1}}, \vec{c_2^{v_2}}\}}{\arg\min} \langle v_1, v_2 \rangle \tag{4.10}$$

We want to create a first orthogonal basis M_1 from u_1 and v_1, and a second orthogonal basis M_2 from u_2 and v_2, so we have:

- $\vec{q}_1$=the projection of $\vec{v}_1$ on the orthogonal plane to $\vec{u}_1$

- $\vec{q}_2$=the projection of $\vec{v}_2$ on the orthogonal plane to $\vec{u}_2$

$$Proj_{orth\_Plane_{\vec{u}_1}}(\vec{v}_1) = \vec{v}_1 - \frac{(\vec{v}_1.\vec{u}_1)}{|\vec{u}_1|^2}\vec{u}_1 \tag{4.11}$$

$$Proj_{orth\_Plane_{\vec{u}_2}}(\vec{v}_2) = \vec{v}_2 - \frac{(\vec{v}_2.\vec{u}_2)}{|\vec{u}_2|^2}\vec{u}_2 \tag{4.12}$$

$$\vec{l_1} = \vec{u}_1 \times \vec{q}_1 \tag{4.13}$$

$$\vec{l_2} = \vec{u}_2 \times \vec{q}_2 \tag{4.14}$$

we have:

$$M_1 = \begin{bmatrix} u_1(x) & q_1(x) & l_1(x) \\ u_1(y) & q_1(y) & l_1(y) \\ u_1(z) & q_1(z) & l_1(z) \end{bmatrix} \tag{4.15}$$

and

$$M_2 = \begin{bmatrix} u_2(x) & q_2(x) & l_2(x) \\ u_2(y) & q_2(y) & l_2(y) \\ u_2(z) & q_2(z) & l_2(z) \end{bmatrix} \tag{4.16}$$

We can calculate the rotation as the basis transfer matrix between M_1 and M_2:

$$M_2 = RM_1 \tag{4.17}$$

$$R = M_2 M_1^{-1} \tag{4.18}$$

We consider the matrix R as the optimal rotation that aligns the three directional Clusters of two scans.

4.4.4 Translation estimation

We have defined a hypothesis as three pairs of planes. Each pair is extracted from two matched clusters $(C_i \in scanA, C_j \in scanB)$.

For each hypothesis we have:

$pair_l = \{Plane_{i,l}(N_{i,l}, d_{i,l}), Plane_{j,l}(N_{j,l}, d_{j,l})\}$.

where l represents the pair index in the hypothesis and i,j represent the indices of the two matched clusters. For each pair we have the following properties:

$$N_l = RN_{i,l} \tag{4.19}$$

$$N_l T = d_{i,l} - d_{j,l} \tag{4.20}$$

therefore according to the two equations above and the definition of a hypothesis, we can deduce the following linear system:

$$\begin{bmatrix} N_1(x) & N_1(y) & N_1(z) \\ N_2(x) & N_2(y) & N_2(z) \\ N_3(x) & N_3(y) & N_3(z) \end{bmatrix} \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} = \begin{bmatrix} d_{i,1} - d_{j,1} \\ d_{i,2} - d_{j,2} \\ d_{i,3} - d_{j,3} \end{bmatrix} \tag{4.21}$$

by solving this linear system, we can find the translation that aligns the three pairs of planes.

4.4.5 Evaluation and discussion

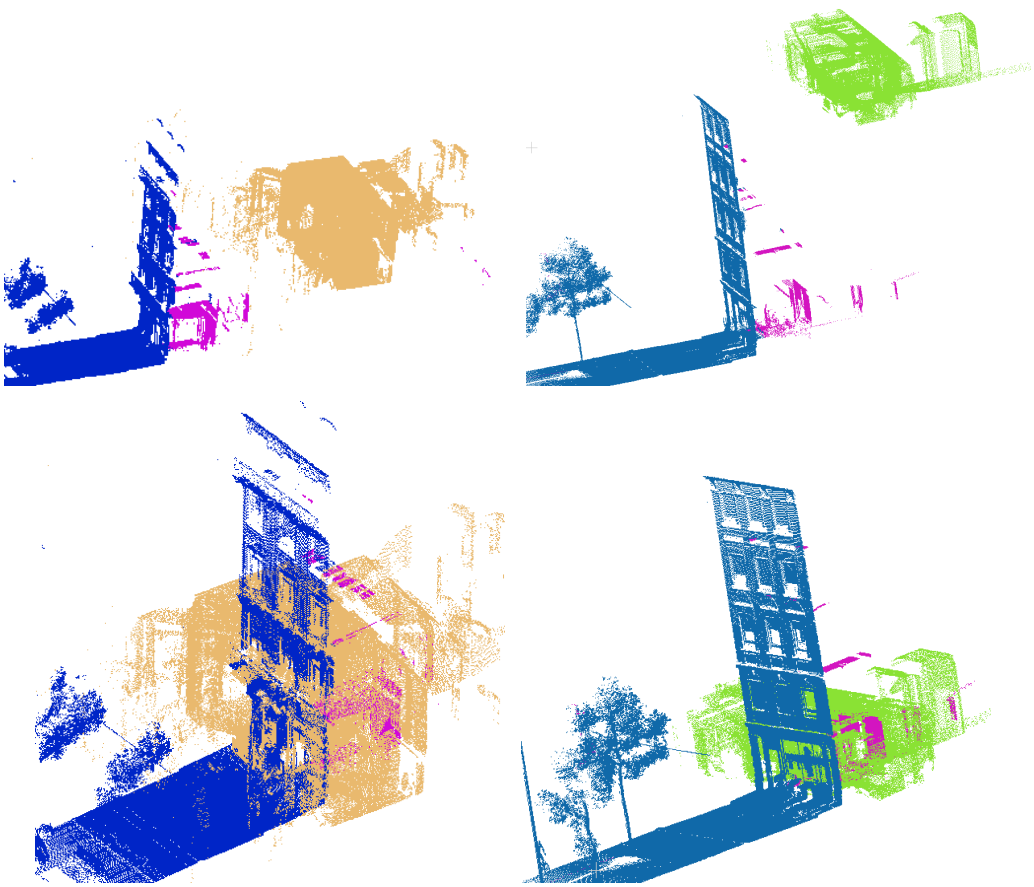


Figure 4.9: Result of the test of our algorithm on two pairs of indoor/outdoor scans, Top: the initial position of each pair, Bottom: The position of each pair after registration, pink points represent the indoor points detected from the outdoor scans.

Polygon MSAC Our proposed method has been tested on real data to evaluate its effectiveness. The proposed algorithm works without any constraints on the initial

Initial error		Final error		Time
Metric error	Angular error	Metric error	Angular error	
0.099 m	0.001°	0.00793 m	0.0008°	263 s
0.44 m	0.528°	0.0213 m	0.06°	244s
1.27 m	3°	0.0241 m	0.043°	269 s
4.15 m	12°	0.0402 m	2°	294 s

Table 4.1: Result of indoor/indoor registration tests

position of the two scans, unlike iterative methods which require the correct estimation of the initial position to be able to converge to a global solution. The key step of our algorithm is the estimation of planar polygons which was carried out using two methods depending on the nature of the data.

The first evaluation of our algorithm was carried out on two interior scans from which a perfect registration was done manually. Each input scan was subsampled with factor of 10, yielding around 260,000 points. Starting from this perfect position, we have altered one of the two scans with an initial translation error ranging from a few centimeters to a few meters and an initial rotation error ranging from a few degrees to a few tens of degrees as shown in Table 4.1. The chosen setting allowed us to estimate 22 planes from the first scan, and 23 planes from the second scan. The obtained results show that our algorithm is able to register the two scans within a reasonable calculation time regardless of the initial error. Afterwards, we performed the registration of indoor and outdoor scans. As we do not know the ground truth we only considered the visual results. We consider that the obtained results demonstrate our algorithm's ability to efficiently register indoor and outdoor data as shown in Figure 4.9.

Achieving precise results requires fine-tuning of the algorithm's parameters. The iteration number of MSAC must be calculated according to the number of points of the scan to ensure the robustness of the algorithm. The number of estimated planes depends on the inlier threshold and the minimum size of a planar region, whereas the number of extracted polygons depends on the value of α. If we estimate more planes, we can generate more hypotheses and use more polygons for the evaluation of each hypothesis, and therefore we can find a more precise result.

4.5 Openings based registration

The topic of building reconstruction draws increasing interest and has various applications such as urban planning, virtual tourism. BIM is also becoming a *de facto* standard to handle the whole life cycle of buildings and 3D city models are central to manage smart cities and in particular to realize environmental simulations. Combining indoor and outdoor modeling of buildings remains a challenging task. The first issue of such modeling is to precisely register the indoor and outdoor data sources. The lack of overlap between indoor and outdoor data is the most encountered obstacle, more so when both data sets are acquired separately and using different types of sensors. Openings are the most obvious common entity to link the inside and outside data. As such, it can help the registration of indoor and outdoor point clouds. So it must be automatically, accurately and efficiently extracted. Therefore, in order to improve the indoor/outdoor registration, we integrate the openings in a registration framework after having extracted them using the method proposed in Section 3.7.

Registration is usually performed by matching features. In the present work, the features are the two sets of vertical openings rectangles detected from indoor and outdoor scans. Selection of correspondences is a crucial step for a successful registration because a bad choice can lead to a bad estimate of the optimal transformation. In our case, the openings are not characteristic enough to match them robustly independently. As an opening is defined by a rectangular shape composed of four segments, two horizontal and two vertical, we can write our registration problem as a minimization of a global robust energy between two segment sets. To achieve this minimization with high accuracy and robustness, we must select the right matches and choose the good optimization algorithm.

4.5.1 Global robust energy between segment sets

We propose to define a robust energy between two sets of 3D segments in a way that minimizing this energy will favor important overlaps between segments and small distances over these overlaps while being robust to outliers, which is to be expected as it is possible that openings extracted form one data set will have no counterpart

in the other. This will implicitly also minimize angles between the segments as an important angle with an important overlap implies a large distance. Given two sets of 3D segments $[A_i^1 B_i^1]$ and $[A_j^2 B_j^2]$ with center points $G_i^k = (A_i^1 + B_i^1)/2$ and direction vectors $\mathbf{d}_i^k = \overrightarrow{A_i^k B_i^k}/||\overrightarrow{A_i^k B_i^k}||$, we start by defining a relative overlap between two segments over the projection p on the bisector of the two segments:

$$overlap([A_i^1 B_i^1], [A_j^2 B_j^2]) = \frac{|p([A_i^1 B_i^1]) \cap p([A_j^2 B_j^2])|}{min(|p([A_i^1 B_i^1])|, |p([A_j^2 B_j^2])|)}$$

We define this bisector as the line passing through the closest point $P_{i,j}$ to $(A_i^1 B_i^1)$ and $(A_j^2 B_j^2)$ and oriented by the bisector $\mathbf{v}_{i,j}$ of $\mathbf{d}_i^1$ and $\mathbf{d}_j^2$. Calling:

$$[\mathbf{v}]_\times = \begin{pmatrix} 0 & -v_z & v_y \\ v_z & 0 & -v_x \\ -v_y & v_x & 0 \end{pmatrix} \text{ such that } [\mathbf{v}]_\times \mathbf{u} = \mathbf{v} \times \mathbf{u} \tag{4.22}$$

we have:

$$P_{i,j} = ([\mathbf{d}_i^1]_\times^t [\mathbf{d}_i^1]_\times + [\mathbf{d}_j^2]_\times^t [\mathbf{d}_j^2]_\times)^{-1} ([\mathbf{d}_i^1]_\times^t [\mathbf{d}_i^1]_\times A_i^1 + [\mathbf{d}_j^2]_\times^t [\mathbf{d}_j^2]_\times A_j^2) \tag{4.23}$$

The projection on the bisector is:

$$p(P) = P_{i,j} + (\overrightarrow{P_{i,j}P}.\mathbf{v}_{i,j})\mathbf{v}_{i,j} \tag{4.24}$$

and the projected points have a linear coordinate:

$$c(P) = \overrightarrow{P_{i,j}P}.\mathbf{v}_{i,j} \tag{4.25}$$

The length of the projected segment is then defined as

$$|p([A_i^1 B_i^1])| = max(c(A_i^1), c(B_i^1)) - min(c(A_i^1), c(B_i^1)) \tag{4.26}$$

and the overlap length of the projected segments as:

$$
\begin{aligned}
&|p([A_i^1 B_i^1]) \cap p([A_j^2 B_j^2])| = \\
&\min(\max(c(A_i^1), c(B_i^1)), \max(c(A_j^2), c(B_j^2))) \\
&- \max(\min(c(A_i^1), c(B_i^1)), \min(c(A_j^2), c(B_j^2)))
\end{aligned} \tag{4.27}
$$

We then define an energy between a segment $L_1 \in \mathcal{L}_1$ and a segment set $\mathcal{L}_2$:

$$
E^{d_{thr}}(L_1, \mathcal{L}_2) = |L_1^2| d_{thr}^2 - \tag{4.28}
$$

$$
\sum_{L_2 \in \mathcal{L}_2} |L_1^2 \cap L_2^1| \max(0, d_{thr}^2 - dist(L_1, L_2)^2) \tag{4.29}
$$

where d_{thr} is a distance threshold above which a line is just considered as an outlier and $dist$ is a metric distance between segments, for instance:

$$
\begin{aligned}
&dist([A_i^1 B_i^1], [A_j^2 B_j^2]) = \\
&\tfrac{1}{4} \left(dist(A_i^1, [A_j^2 B_j^2]) + dist(B_i^1, [A_j^2 B_j^2]) + dist(A_j^2, [A_i^1 B_i^1]) + dist(B_j^2, [A_i^1 B_i^1]) \right)
\end{aligned} \tag{4.30}
$$

For each line pair, this denergy is 0 if L_1 is completely overlapped by segments of $\mathcal{L}_2$ with 0 distance over these overlaps. Finally we can write our final symmetrized robust energy between 3D segments sets as:

$$
E_S(\mathcal{L}_1, \mathcal{L}_2) = \sum_{L_1 \in \mathcal{L}_1} E^{d_{thr}}(L_1, \mathcal{L}_2) + \sum_{L_2 \in \mathcal{L}_2} E^{d_{thr}}(L_2, \mathcal{L}_1) \tag{4.31}
$$

4.5.2 RANSAC optimization

RANSAC has proven its robustness and efficiency as an optimization algorithm in several applications. In this part, we describe a new RANSAC based on two matching strategies. The first consists in matching the vertical planes representing the walls of the two scans whereas the second consists in matching the selected segments in the matched walls. The advantage of the integration of these two matching strategies is to simplify the selection of the two pairs of segments and optimize the calculation time of the algorithm by reducing the number of iterations. At each RANSAC iteration:

- We randomly select a wall plane from each scan.
- We randomly select one vertical segment and one horizontal segment in each selected wall plane.
- We associate the vertical segments and the horizontal segments between the two selected sets.
- We compute the transform (rotation and translation) that best aligns the matched 3D segments using the method of Section 4.5.3.
- For this transform, we estimate the global robust energy between all segments of the two sets using (4.31).

The final registration is given by the transformation that minimizes the global robust energy.

4.5.3 Transform estimation

Rotation estimation

Once two pairs of segments $\{v_i, v_j\}$ (vertical) and $\{h_i, h_j\}$ (horizontal) are associated, we estimate the rotation that best aligns the corresponding two 3D lines. Let us call d_v^i, d_v^j, d_h^i and d_h^j the unit norm vectors of v_i, v_j, h_i and h_j. We start by creating orthonormal bases $\mathcal{O}^i = (\mathbf{x}^i, \mathbf{y}^i, \mathbf{z}^i)$, where:

$$\mathbf{x}^i = \mathbf{d_v^i} \quad \mathbf{y}^i = \frac{\mathbf{d_h^i} - (\mathbf{d_h^i} \cdot \mathbf{x}^i)\mathbf{x}^i}{||\mathbf{d_h^i} - (\mathbf{d_h^i} \cdot \mathbf{x}^i)\mathbf{x}^i||} \quad \mathbf{z}^i = \mathbf{x}^i \times \mathbf{y}^i. \tag{4.32}$$

We then compute the rotation R that aligns the associated segments as the base change matrix between $\mathcal{O}^i$ and $\mathcal{O}^j$:

$$R = \mathcal{O}^j \mathcal{O}^{i^{-1}}. \tag{4.33}$$

Translation estimation

To estimate the translation, we start by defining the point to line distance:

$$dist(\mathbf{p}, L = \mathbf{a} + \mathbf{d}t) = \frac{||\mathbf{d} \wedge (\mathbf{a} - \mathbf{p})||}{||\mathbf{d}||} = [\mathbf{d}]_\times(\mathbf{a} - \mathbf{p}) \tag{4.34}$$

assuming that $\mathbf{d}$ is normalized, and again calling $[\mathbf{d}]_\times$ the matrix of the cross product with $\mathbf{d}$. We look for the translation $\mathbf{t}$ that minimizes:

$$MSACepsilon = \sum_i ||[\mathbf{d}_i]_\times(\mathbf{a}_i - (\mathbf{p}_i + \mathbf{t}))||^2 \tag{4.35}$$

The minimum is reached where the gradient vanishes:

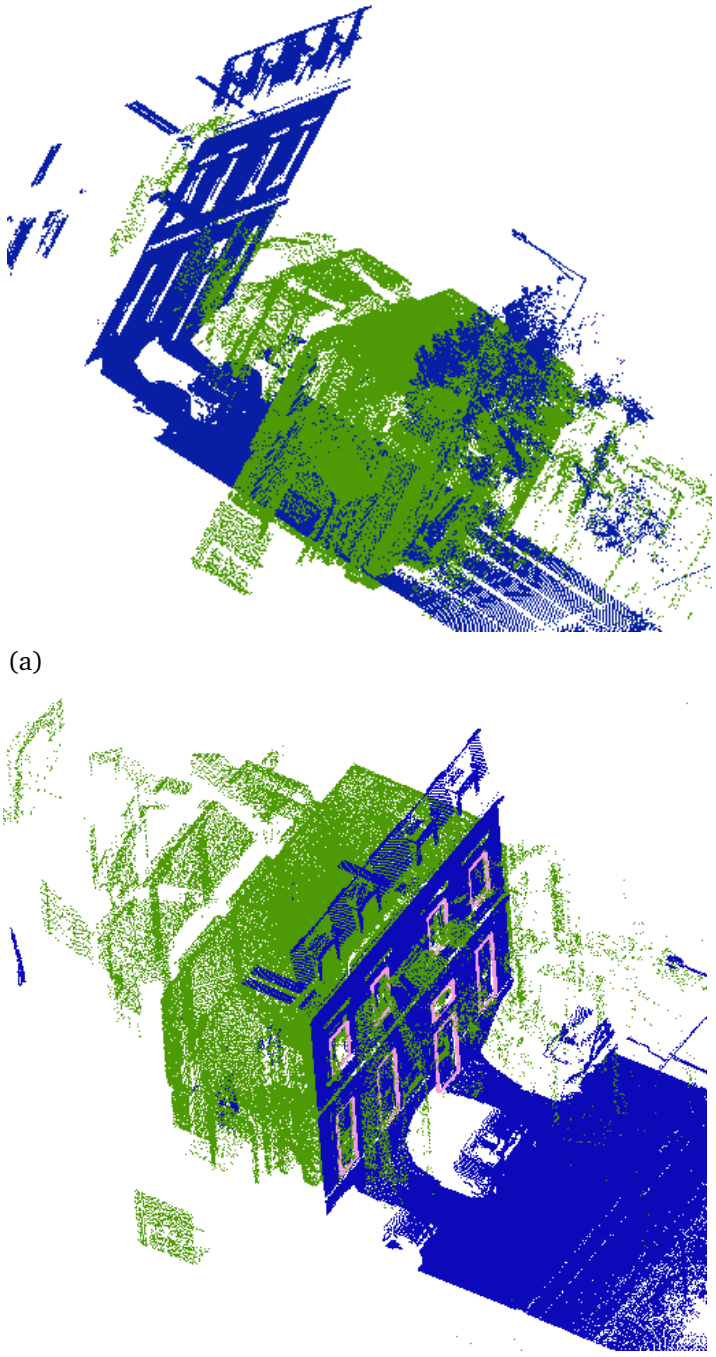
$$\nabla_\mathbf{t}\epsilon(\mathbf{t}) = -2\sum_i [\mathbf{d}_i]_\times^t [\mathbf{d}_i]_\times(\mathbf{a}_i - (\mathbf{p}_i + \mathbf{t})) = 0. \tag{4.36}$$

Calling:

$$\mathbf{w} = \sum_i [\mathbf{d}_i]_\times^t [\mathbf{d}_i]_\times(\mathbf{p}_i - \mathbf{a}_i) \quad M = -\sum_i [\mathbf{d}_i]_\times^t [\mathbf{d}_i]_\times \tag{4.37}$$

We get as translation vector:

$$\mathbf{t} = M^{-1}\mathbf{w}. \tag{4.38}$$



(a)

(b)

Figure 4.10: Indoor/Outdoor registration results: (a) Position of the two scans before registration, (b) the position after registration (Green: indoor scan, Blue: outdoor scan, Pink: outline of outdoor openings).

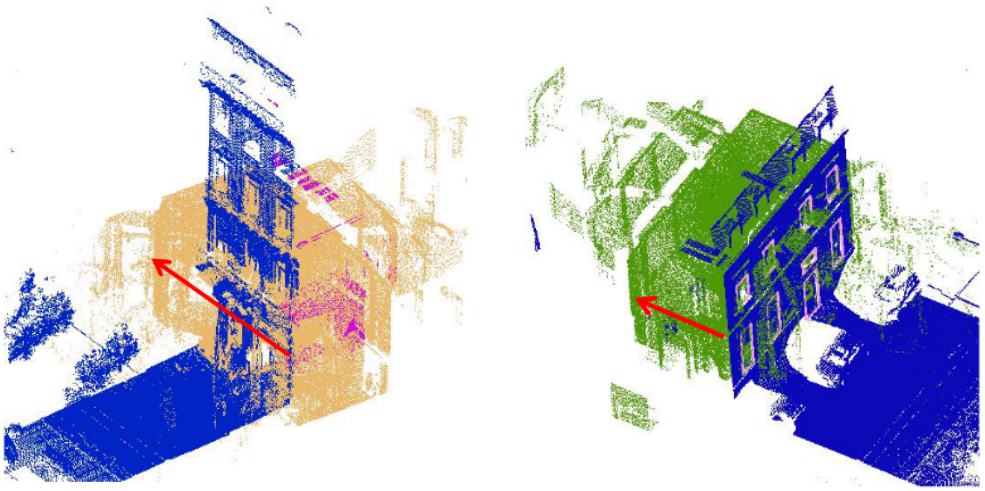
4.5.4 Evaluation and discussion

In this section, our contribution consists in proposing an openings based registration algorithm to address the challenging problem of indoor/outdoor registration with a unified formalism. Our proposed algorithm works without any constraints on the initial position of the two scans, unlike iterative methods that require a good ap-

proximation of the initial transformation to be able to converge towards the correct solution. The key step of our algorithm is the detection of openings which was carried out using an automatic method applicable in indoor and outdoor data. The obtained results have proven the performance of our algorithm to register indoor/outdoor data whatever the initial position and orientation as shown in Figure 4.10. We defined a global robust energy between two segments sets and we proposed a robust approach to minimize this energy using the RANSAC pardigm.

4.6 Hybrid solution

In Sections 4.4 and 4.5, we have proposed two efficient algorithms to deal with indoor/outdoor registration problem. The two proposed algorithms have proven their performance to solve this challenging problem, but they have some limitations. For the planar polygons based solution, we have an uncertainty in the horizontal direction parallel to the facade. For the openings based solution, we have an uncertainty in the direction orthogonal to the facade as shown in Figure 4.11.



(a) (b)

Figure 4.11: Uncertainties of the two proposed solutions, (a) planar polygons based solution: an uncertainty in the horizontal direction parallel to the facade, (b) openings based solution, an uncertainty in the direction orthogonal to the facade.

In order to remove these uncertainties, we have tried to combine the two solutions. This combination allows us to generate a hybrid solution able to perform the indoor/outdoor registration with great precision and exceeding the limits of the two previous solutions.

4.6.1 Valid hybrid associations

In this approach, each scan (indoor scan or outdoor scan) is simplified by two types of geometric primitives: openings and planar polygons. In order to accelerate and simplify the matching of these components between the two scans, we propose to start by efficiently associating the supporting planes of polygons and the openings of the same scan. This is done using the Algorithm 2.

Algorithm 2 Associations Creation

Input: a set of supporting planes of polygons $S_1 = \{P_p\}$ and a set of supporting planes of openings $S_2 = \{P_o\}$ extracted form the same scan.
Output: a list of associations of hybrid supporting planes $L = \{A(P_o, P_p)\}$.
for each supporting plane P_o **do**
  for For each supporting plane P_p **do**
    if $Angle({P_o}^{\perp}, {P_p}^{\perp}) < \epsilon$ **then**
      Create an association $A(P_o, P_p)$ and add it to L.
    end if
  end for
end for

4.6.2 Global hybrid energy

We propose to define a hybrid energy between two hybrid association sets. Each association sets corresponds to a scan. Each association is formed by a supporting plane of a polygon and a supporting plane of an opening extracted from the same scan. As an opening is defined by a rectangular shape composed of four segments, two horizontal and two vertical ones, the minimization of the hybrid energy must favor important overlaps both between polygons and between segments and small distances over these overlaps. For robustness, we need a distance threshold d_{thr} above which a polygon or a 3D segment is considered as an outlier. This hybrid energy is inspired by the equations (4.9) and (4.31).

$$
\begin{aligned}
&E^{d_{thr}}(A_i(P_i, L_i) \in Scan_1, Scan_2) \approx \\
&d_{thr}^2 - \sum_{A_j \in Scan_2} \sum_{P_j \in A_j} \frac{|P_i' \cap P_j'|}{\min\left(|P_i'|, |P_j'|\right)} max(0, d_{thr}^2 - Distance(P_i, P_j)^2) + \\
&|L_i| d_{thr}^2 - \sum_{A_j \in Scan_2} \sum_{L_j \in A_j} |L_i' \cap L_j'| \max(0, d_{thr}^2 - Distance(L_i, L_j)^2)
\end{aligned}
\tag{4.39}
$$

where:

- $|.|$ represents the area for the polygons and the length for the 3D segments.

We see that the energy is null when P_j completely overlaps by P_i (relative overlap is 1) with a distance $Distance(P_i, P_j) = 0$ and L_j completely overlaps L_i with a distance $Distance(L_i, L_j) = 0$.

Finally we can write our final symmetric hybrid energy as:

$$E_h(Scan_1, Scan_2) = \sum_{A_i(P_i,L_i)\in Scan_1} E^{d_{thr}}(A_i(P_i, L_i), Scan_2) + \sum_{A_j(P_j,L_j)\in Scan_2} E^{d_{thr}}(A_j(P_j, L_j), Scan_1) \quad (4.40)$$

4.6.3 RANSAC based optimization of the hybrid energy

To carry out the registration of the two association sets, we now have to find the rotation and the translation that minimize the hybrid energy (4.40). In this section, we describe a new adaptation of RANSAC based on two matching steps. The first consists in matching the valid associations between the two scans whereas the second consists in matching the selected segments and polygons in the matched associations. The advantage of the integration of these two matching steps is to simplify the selection of the two pairs of segments and the pair of polygons, necessary to calculate the transformation at each iteration, and optimize the calculation time of the algorithm by reducing the number of iterations.

At each RANSAC iteration:

- Randomly select an association from each scan ($A_i(P_{o_i}, P_{p_i}) \in Scan_1$ and $A_j(P_{o_j}, P_{p_j}) \in Scan_2$).
- Make the two planes (P_{o_i}, P_{o_j}) coincident by applying a rotation R around their line of intersection, this rotation can be computed using the method of Section 4.6.4..
- Randomly select a corner (the intersection point of a horizontal segment and a vertical segment) from the associated openings of P_{o_i} and P_{o_j}.
- Compute the translation in the plane between the two selected corners (we have only two degrees of freedom).

- Find the missing degree of freedom of the translation by computing the distance between the associated planar polygons of P_{p_i} and P_{p_j}.

Finally, keep the sampled transformation (rotation + translation) that has the minimum hybrid energy (4.40).

4.6.4 Rotation around intersection line

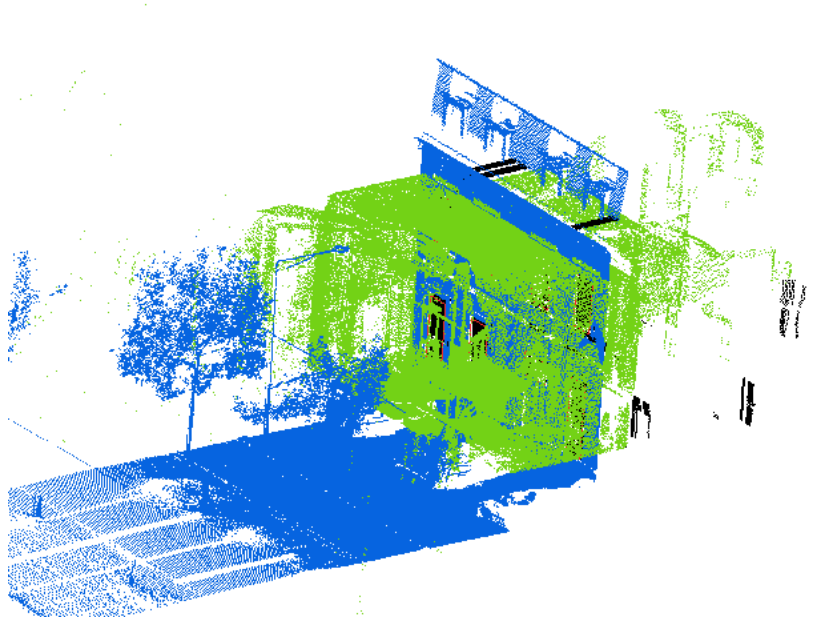
We compute the rotation around the intersection lines of two planes $\mathbf{P}_1$,$\mathbf{P}_2$ as the following:

- N_1 is the normal vector of $\mathbf{P}_1$.
- N_2 is the normal vector of $\mathbf{P}_2$.
- U is the unit vector of the intersection line computed according the the equation (4.41).

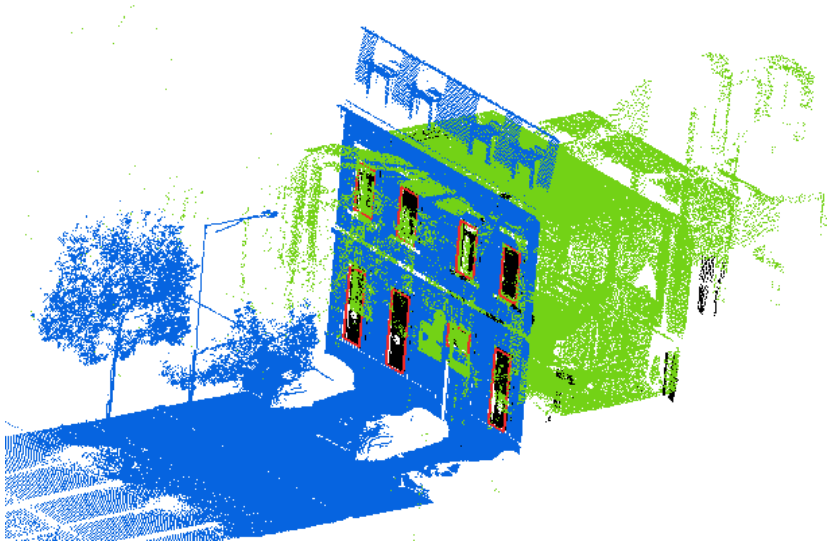
$$U = N_1 \wedge N_2 \tag{4.41}$$

As N_1 and N_2 are normalized, the rotation angle will be:

$$\alpha = \arccos(N_1 \cdot N_2). \tag{4.42}$$



(a)



(b)

Figure 4.12: Indoor/Outdoor registration results by combining polygons and openings:(a) the result of the first step by considering only the openings, (b) the final result by adding the polygons translation , indoor scan (green points), outdoor scan(blue points), indoor points of the outdoor scan (black points), openings outlines of the outdoor scan (red lines).

4.7 Conclusion

In this chapter, we have addressed the indoor /outdoor registration problem which represents a major challenge for Indoor/Outdoor building modeling. We have exploited the geometric and physical properties of our data to propose potential solutions for this challenging problem. As we worked in a building environment, where

most objects are represented by a planar surface that can be simplified by planar polygons, we started by proposing a planar polygons based registration algorithm. This first algorithm matches planar polygons by clustering polygons according to their normal direction, then by their offset in the normal direction. It uses this clustering to find possible polygon correspondences (hypotheses) and estimate the optimal transformation for each hypothesis. The best transformation is selected according to a global robust energy. An opening in the façade is the unique common entity that can be seen from inside and outside. It can help the registration of indoor and outdoor point clouds. For this reason, we have integrated openings in a registration framework in order to propose a second solution. As an opening is defined by a rectangular shape composed of four segments, two horizontal and two vertical, we preferred to write our registration problem as a minimization of a global robust energy between two segment sets. The two proposed solutions are very efficient for realizing the indoor/outdoor registration. But they have some limitations. In order to overcome these limitations and benefit from the advantages of the two solutions, we have tried to combine them. This combination allows us to generate a hybrid solution able of performing indoor/outdoor registration with good qualitative precision.

Chapter 5

Image/LiDAR data Registration

5.1 Introduction

The volume of datasets acquired by optical and LiDAR systems is rapidly increasing due to demand from a variety of applications including remote sensing, 3D mapping and autonomous driving. 3D scene analysis and reconstruction from image and LiDAR is an active research area in computer vision. On the one hand, the LiDAR data provides highly accurate and robust surface information. On the other hand, the image provides high resolution details and rich spectral information (Kim et al., 2006) but the 3D geometry estimated from dense matching is less robust and accurate, in particular in homogeneous, specular or repetitive regions and near depth discontinuities. Hence, integrating data from these two sources can lead to a more robust and complete semantic segmentation and reconstruction of 3D scenes. Following a classical methodology (Kumar Mishra and Zhang, 2012), the problem of image/LiDAR data registration is decomposed in three main steps: (i) Feature extraction, (ii) Feature matching, (iii) Transformation model estimation.

5.2 State of the art

There is a considerable amount of prior work on image/LiDAR registration. We start by classifying some existing methods according to their types and attributes, citing

the advantages and limitations of each category.

Keypoint based methods

Corners are the most used key points for image/image registration, which can be extended to image/LiDAR registration (Ding et al., 2008). The uniqueness and high precision in the localization are the strongest properties of corners (Kumar Mishra and Zhang, 2012). However, corners are not always easily matched between LiDAR and image.

Linear feature based methods

Straight lines are the most used feature for image/LiDAR registration because they can be automatically, accurately and efficiently extracted from the LiDAR and image data (Habib et al., 2006), especially if we have a set of overlapping images (Deng et al., 2008).

Structural feature based methods

High level structural features such as rectangles can help increasing the robustness of both detection and matching steps for image/LiDAR registration. These structural features can be extracted from both data sets as connected segments (Liu and Stamos, 2012). This kind of features can be efficiently used to estimate camera translation. Camera rotation can be estimated using at least two vanishing points. The major limitation of these methods consists in the dependence on either the strong presence of parallel lines to infer vanishing points or availability of feature pair correspondences (Wang and Ferrie, 2015). Moreover, these methods are not effective when applied to aerial data.

Mutual information based methods

Statistical and information theoretic methods have demonstrated excellent performance for a wide variety of 2D-2D and 2D-3D registration applications. Mutual information, a central concept of information theory, consists in measuring the statistical correlation of two random variables, which is a measure of the amount of information one random variable contains about the other (Cover, 1999). Recent

methods use Mutual Information (MI) as a statistical metric to register image and LiDAR data (Lu et al., 2019). A mutual information (MI) approach was proposed in (Mastin et al., 2009). This method performs the registration by seeking the camera matrix that maximizes the MI between the distribution of image features and projected LiDAR features. The authors of (Wang and Ferrie, 2015) search for the optimal camera pose through maximizing the MI between camera image and LiDAR attribues (LiDAR intensity image, LiDAR elevation image) (Wang and Ferrie, 2015). The major limitation of these methods lies in the decreasing registration accuracy when only LiDAR elevation image is used. An hybrid online calibration method for a laser scanner mounted on a mobile platform also equipped with a imaging system has been proposed in (Miled et al., 2016). The inaccurate transformation between camera and laser scanner has been optimized by maximizing the mutual information (MI) between these two kinds of data resulting in an accurate online calibration.

Frequency based registration

The most popular frequency based method is phase correlation. Due to differences in data characteristics, these frequency based methods cannot be applied directly to register images and LiDAR data. (Zhu et al., 2020) has proposed a new registration method to deal with the problem of aerial image and LiDAR registration. This method is based on structural features and 3D phase correlation, in order to address significant geometric distortions and nonlinear intensity differences between the aerial and LiDAR intensity images. The 3D phase correlation is used to detect control points (CPs) between aerial images and LiDAR data in the frequency domain, that will be used to correct the exterior orientation elements. The main limitation of this method lies in the use of the intensity images of LiDAR only, which contains less information about the topography as compared to LiDAR elevation images (Kumar Mishra and Zhang, 2012).

According to this summary comparison, we advocate that the best methods to apply for image/LiDAR registration in urban environments are mutual information and straight line based methods. We have chosen to develop a straight line based algorithm for the following reasons:

- In building environment, most objects are composed of planar surfaces delimited by straight lines.
- On our scenes of interest, we have both overlapping images and LiDAR scans.
- There are efficient algorithms in the literature to automatically and precisely extract these features from both images and LiDAR data.

5.3 Overview and contributions

In a man made environment, we advocate that the best feature to use for image/LiDAR data registration is the straight line segment. Straight line segments can be reliably, accurately and automatically extracted from both LiDAR and image data under sufficient image overlap. They aggregate more information than points therefore less sensitive to more dense noise and more frequent than more complex primitives (rectangles).
In this chapter, we propose a new method for image/LiDAR registration, that consists in:

- Extracting 3D segments from the two datasets.
- Defining a global robust energy between the two segments sets that can be minimized by RANSAC or simulated annealing.

RANSAC works by matching pairs of segments. The Figure 5.1 shows details of our pipeline .

5.4 3D Segment Extraction

5.4.1 3D Line Segment Extraction from LiDAR Data

3D line segment detection from LiDAR data is a crucial step in our registration procedure. We have proceeded to the detection of these features according to the method detailed in Section 3.5.

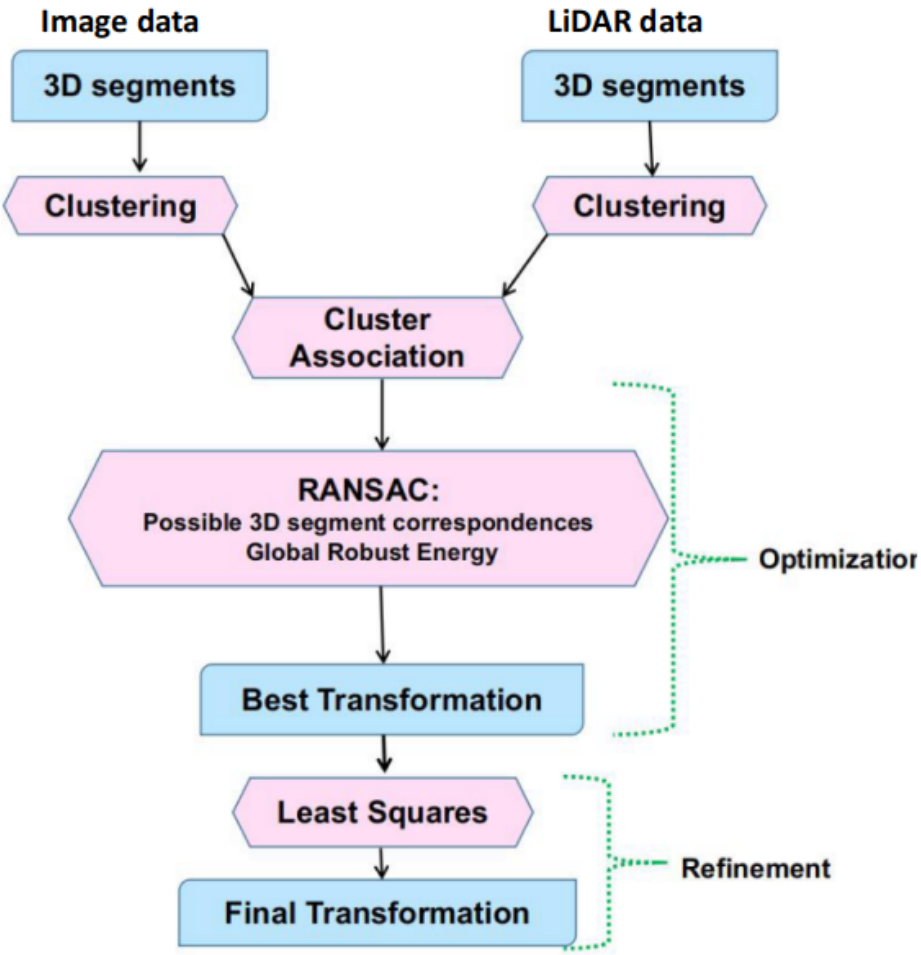


Figure 5.1: Pipeline details of our proposed solution for image/LiDAR data registration problem.

5.4.2 3D line segment extraction from image data

3D line segments reconstruction from 2D images has received continued attention in the photogrammetry and computer vision literature. The first proposed methods used pairwise reconstruction where only two overlapping images are required to perform the reconstruction. Other methods have recently been proposed which use multi-view reconstruction where various overlapping images are used and the camera pose can be obtained using an SfM (Structure from Motion) pipeline. In this work, we chose to work with the method proposed in Section 3.6.

5.5 3D segments based registration

Registration is usually performed either based on a global energy between the two data sets to register, or by matching features. In our case, the line segments are

not characteristic enough to match them robustly independently. This is why we preferred to define a global robust energy between two segment sets and propose a robust approach to minimize this energy. We propose to work with the energy defined in Section 4.5.1. It is a robust energy between two 3D segment sets in a way that minimizing it will favor important overlaps between segments and small distances over these overlaps while being robust to outliers, as many lines extracted form one data set will have no counterpart in the other. This will implicitly also minimize angles between the segments as an important angle with an important overlap implies a large distance.

5.5.1 3D segments directional clustering

In order to simplify the minimization of this energy between segments sets, we start by clustering the segments (by direction). Then, we define the rotation between the segments sets by matching clusters and not individual segments in order to be more robust and precise. This is done using our proposed greedy algorithm described in Algorithm 3.

Algorithm 3 Greedy direction clustering

1: Input: Set of segments L, each segment $L_i = [A_iB_i]$ has a director vector $\mathbf{v}_i = \overrightarrow{A_iB_i}$, a length $len_i = ||\mathbf{v}_i||$ and a unit direction $\mathbf{d}_i = \mathbf{v}_i/len_i$.

2: Initialize an empty set of 3D segment clusters $\mathcal{C}$. We will call direction of a cluster C the weighted mean of the directions of the 3D segments:

$$\mathbf{d}(C) = \frac{\sum_{L_i \in C} sign(\mathbf{v}_i \cdot \mathbf{v}_1)\mathbf{v}_i}{||\sum_{L_i \in C} \mathbf{v}_i||}$$

3: For each segment L_i in descending order of length (from longest to shortest):
- If $\mathcal{C} = \emptyset$ or $\max_{C \in \mathcal{C}} \mathbf{d}_i \cdot \mathbf{d}(C) < \cos\epsilon$, create a new cluster $C = \{L_i\}$ and add it to $\mathcal{C}$. Note that this is equivalent to finding if L_i has an angle smaller than ϵ with an existing cluster direction but this formulation with dot products is faster to compute.
- Else add L_i to the cluster $\arg\max \mathbf{d}_i \cdot \mathbf{d}(C)$.

5.5.2 Valid cluster associations

To define a rotation between the two segment sets, it is sufficient to associate two pairs of segments from the two sets. However this is not very precise, so we propose to associate the segment clusters defined above. We propose two possible sets of

valid associations depending on whether we have a vertical cluster for each scan.

LiDAR vertical cluster selection

We assume that all LiDAR scans are vertically oriented (this is the case for most modern scanners) and select the cluster with the smallest angle with the Z-axis $(0, 0, 1)$ as being the vertical cluster.

Image vertical cluster selection

We assume that the images are upright (the real world vertical projects as a nearly vertical 2D vector in all images) and select the vertical cluster of the line cloud reconstructed from the images according to Algorithm 4.

Algorithm 4 Image vertical cluster

for each cluster i **do**
  for each 3D line j in cluster i **do**
    for each 2D image im associated with line j **do**
      Compute S_{im} the verticality score weighted by the length of the 2D segment $l_i = [a_i b_i]$ in the image im that has been used for the reconstruction of line j using the method proposed in Section 5.5.2
    end for
    Compute av_j: the average verticality score for all images associated with line j
  end for
  Compute AV_i: the average verticality score for all 3D lines in cluster i
end for
Select the cluster which has the smallest score as the vertical cluster

Verticality score calculation

We have an image im which contains 2D segment $l_i = [a_i b_i]$. This 2D segment l_i has a director vector $v_i = \overrightarrow{a_i b_i}$ and a unit direction $d_i = v_i / ||v_i||$. We compute S_{im} the verticality score weighted by the length of the 2D segment as the angle between d_i and y-axis $(0, 1)$ according to the equation (5.1):

$$im = ||v_i|| \langle d_i, y \rangle \tag{5.1}$$

If the two segment sets $\mathcal{S}_1$ and $\mathcal{S}_2$ (image and/or LiDAR) can be vertically oriented (the assumptions above are true), we obviously associate the vertical cluster of $\mathcal{S}_1$ to the vertical cluster of $\mathcal{S}_2$, and then associate any non vertical cluster of a $\mathcal{S}_1$ with any

non vertical cluster of $\mathcal{S}_2$. In the other case (at least one scan cannot be vertically oriented), we associate any pair of clusters of $\mathcal{S}_1$ to any pair of clusters of $\mathcal{S}_2$. As for each cluster we have two possible direction vectors:$\{\mathbf{d}(C), -\mathbf{d}(C)\}$, we can:
define the variables $s_i^j : i \in [1,2], j \in [1,2], s_i^j \pm 1$

Calling C_1^1, C_2^1 a pair of clusters of $\mathcal{S}_1$ and C_1^2, C_2^2 a pair of clusters of $\mathcal{S}_2$. In the two cases, for each cluster association $As = \{(C_1^1, C_2^1) \leftrightarrow (C_1^2, C_2^2)\}$, we have sixteen possible forms. If one of these sixteen forms satisfies equation (5.2), we consider that this association is valid, we use the retained form to calculate the rotation. We reject the cluster association, if the equation (5.2) is not validated by any form.

$$|\angle(s_1^1\mathbf{d}(C_1^1), s_2^1\mathbf{d}(C_2^1)) - \angle(s_1^2\mathbf{d}(C_1^2), s_2^2\mathbf{d}(C_2^2))| < \epsilon \tag{5.2}$$

5.5.3 Rotation estimation

Once the best form of the valid cluster association $As = \{(s_1^1 C_1^1, s_2^1 C_2^1) \leftrightarrow (s_1^2 C_1^2, s_2^2 C_2^2)\}$ is selected, we want to find the rotation that best aligns the corresponding two pairs of directions. After selecting the directing vectors of two clusters. We start by creating orthonormal bases $\mathcal{O}_i = (\mathbf{x}^i, \mathbf{y}^i, \mathbf{z}^i)$, where:

$$\mathbf{x}^i = s_1^i\mathbf{d}(C_1^i)$$

$$\mathbf{y}^i = \frac{s_2^i\mathbf{d}(C_2^i) - s_2^i\mathbf{d}(C_2^i)\cdot\mathbf{x}^i\mathbf{x}^i}{||s_2^i\mathbf{d}(C_2^i) - s_2^i\mathbf{d}(C_2^i)\cdot\mathbf{x}^i\mathbf{x}^i||}$$

$$\mathbf{z}^i = \mathbf{x}^i \times \mathbf{y}^i \tag{5.3}$$

We then compute the rotation R that aligns the associated clusters as the base change matrix between $\mathcal{O}^1$ and $\mathcal{O}^2$:

$$R = \mathcal{O}_2\mathcal{O}_1^{-1}. \tag{5.4}$$

5.5.4 Optimization of a global robust energy between two segments set

To carry out the registration of the two segment sets, we now have to find the rotation, the scale factor and the translation that minimize a global robust energy be-

tween two segments sets. we have chosen to work with the energy (4.31) which has been defined in the openings based algorithm. We want this minimization to be insensitive to initialization, so we propose two randomized optimizers: Simulated Annealing (SA) and RANSAC.

Simulated annealing optimization

We use simulated annealing to explore the translation and scale parameters. In simulated annealing, a new solution is iteratively computed in the vicinity of the current solution and this new solution is accepted with a certain probability depending on its energy (the global robust energy in our case). However for rotations we want to speed up the process by only exploring valid cluster associations to limit the complexity of the search, see Algorithm 5.

Algorithm 5 Simulated Annealing

Input:
M_0: initial transformation (randomly chosen valid cluster association and randomly chosen scale/translation)
E_0: initial transformation energy
G: Current transformation
T_{max}: initial temperature
T_{min}: final temperature
α: cooling rate
$MaxIter$: maximum number of iterations
$G = M_0$, $E(G) = E_0$
while $T_{max} > T_{min}$ **do**
  for $Iter = 1 \dots$ MaxIter **do**
    With a small probability p_{jump}, randomly select a new valid cluster association and corresponding rotation.
    Randomly sample a new scale/translation close to the current one, and call M_N the resulting transformation.
    Compute $\Delta_E = E(M_N) - E_0$
    if $(\Delta_E < 0)$ or $(random < e^{\frac{-\Delta_E}{T_{max}}})$ **then**
      $M_0 = M_N$
      $E_0 = E(M_N)$
    end if
    if $(E_0 < E(G))$ **then**
      $G = M_0$
      $E(G) = E_0$
    end if
  end for
  $T_{max} = \alpha * T_{max}$
end while

The solution witch has the minimum energy is considered as the optimal solution

RANSAC optimization

The adaptation of RANSAC to valid associations is quite straightforward. At each RANSAC iteration, we randomly select a valid cluster association, then randomly select one 3D segment in each of the associated clusters. If the distance between the supporting lines of two segments is smaller than a threshold d_{min} chosen to be the expected noise level in the corresponding scan, we reject it because matching coplanar segments will lead to degenerate scale estimation. We then compute the rotation based on cluster association using the method of Section 5.5.3 and estimate the scale/translation that best aligns the associated 3D segments.

Finally, we keep the sampled transformation that has the minimum robust energy. Optionally, we can refine this solution by matching all pairs of segments that are close enough in the two scans (with a distance smaller than d_{thr}) and recompute the scale/translation based on all these associations to have a more precise estimation.

Scale/translation estimation

The goal of this step is to estimate the scale/translation that minimizes the distance between matched segments pairs. To do this, we define the point to line distance as

$$dist(\mathbf{p}, L = \mathbf{a} + \mathbf{d}t) = \frac{||(\mathbf{a} - \mathbf{p}) \wedge \mathbf{d}||}{||\mathbf{d}||}. \tag{5.5}$$

Assuming that $\mathbf{d}$ is normalized, this writes

$$dist(\mathbf{p}, L = \mathbf{a} + \mathbf{d}t) = \|[\mathbf{d}]_\times(\mathbf{a} - (s\mathbf{p} + \mathbf{t}))\|. \tag{5.6}$$

Our goal is to find the optimal translation $\mathbf{t}$ and scale s that minimize

$$\epsilon(s, \mathbf{t}) = \sum_i |||[\mathbf{d}]_i \times (\mathbf{a}_i - (s\mathbf{p}_i + \mathbf{t}))||^2. \tag{5.7}$$

The minimum is reached where the gradient vanishes:

$$\nabla_{\mathbf{t}}\epsilon(s,\mathbf{t}) = 2\sum_i [\mathbf{d}_i]_\times^2(\mathbf{a}_i - (s\mathbf{p}_i + \mathbf{t})) = 0, \tag{5.8}$$

$$\nabla_s\epsilon(s,\mathbf{t}) = 2\sum_i \mathbf{p}_i^t[\mathbf{d}_i]_\times^2(\mathbf{a}_i - (s\mathbf{p}_i + \mathbf{t})) = 0. \tag{5.9}$$

Calling:

$$\mathbf{w}_1 = \sum_i [\mathbf{d}_i]_\times^2\mathbf{a}_i \qquad M_1 = \sum_i [\mathbf{d}_i]_\times^2 \qquad M_2 = \sum_i [\mathbf{d}_i]_\times^2\mathbf{p}_i$$

$$\mathbf{w}_2 = \sum_i \mathbf{p}_i^t[\mathbf{d}_i]_\times^2\mathbf{a}_i = -\sum_i (\mathbf{d}_i \times \mathbf{p}_i)\cdot(\mathbf{d}_i \times \mathbf{a}_i)$$
$$M_3 = \sum_i \mathbf{p}_i^t[\mathbf{d}_i]_\times^2\mathbf{p}_i = -\sum_i \|\mathbf{d}_i \times \mathbf{p}_i\|^2$$

we get

$$\begin{bmatrix}\mathbf{t}\\ s\end{bmatrix} = \begin{bmatrix}M_1 & M_2\\ M_2^t & M_3\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{w}_1\\ \mathbf{w}_2\end{bmatrix}. \tag{5.10}$$

5.6 Iterative Closest Line (ICL)

We have integrated the clustering and the global robust energy in the Iterative Closest Line (ICL) paradigm in order to compare it with our algorithm in terms of performance and robustness.

5.6.1 Matching step

We applied a filter based on three criteria to select the correspondences between two 3D segment sets: Angle, distance and overlap.

- Angle: in order to simplify the validation of this criterion, we preferred to select segments belonging to the associated clusters (to avoid choosing a threshold for this criterion);
- Distance: we have defined this distance according to (4.30).
- Overlap: we have defined the overlap between two 3D segments according

to (4.27).

So at each iteration, we select the set of segments pairs which validates the three criteria as the matching set.

5.6.2 Optimization step

At each iteration:

- We have estimated the rotation matrix that aligns the direction vectors of the matched 3D segments.
- We have used the equations of Section 5.5.4 in order to find the scale factor and the translation between the matched segments after having aligned them using the estimated rotation.
- We computed the energy (global robust energy) of the estimated transformation (rotation, scale factor, translation).

Finally, we keep the sampled transformation that has the minimum energy.

5.6.3 Rotation estimation

We have a set of matched lines. Each line L_i has a direction vector $\mathbf{d}_i$. Following (Alshawa, 2007), the rotation matrix can be computed from line directions only by minimizing the following function:

$$Err(R) = \sum_i ||\mathbf{d_i^2} - \mathbf{R}\mathbf{d_i^1}||^2. \tag{5.11}$$

The solution of this function is of closed form. For doing so, the two following means are defined:

$$\mu^1 = \frac{1}{N}\sum_i^N \mathbf{d_i^1} \qquad\qquad \mu^2 = \frac{1}{N}\sum_i^N \mathbf{d_i^2} \tag{5.12}$$

The cross covariance matrix of the two data sets is:

$$A = \sum_i^N (\mathbf{d_i^2}\mathbf{d_i^1}^T - \mu^2\mu^{1T}) \tag{5.13}$$

By using SVD (singular value decomposition) of A, we can find R as:

$$R = \mathbf{U}\mathbf{V}^T \tag{5.14}$$

5.6.4 Limitations

In practice, finding appropriate thresholds for the distance and overlap is tedious and often leads to multiple or no matchings. In addition, ICL requires a good approximation of the initial transformation to be able to converge towards the right solution. Our RANSAC based approach is the best solution for the registration of two 3D segment sets. The selection of the best transformation as the one which optimizes the global robust energy between all the extracted 3D segments makes our solution the most consistent with all the information of the datasets to be registered.

5.7 Evaluation and discussion

The evaluation of our approach was carried out in two steps using two data types:

5.7.1 Evaluation on realistic data

The first evaluation of our approach was carried out on realistic data. We tried to simulate a realistic setting with access to a perfect ground truth by generating of two copies of a real LiDAR line cloud. The first copy has been generated by removing 25% of the lines and adding noise. The second copy has been generated by removing 33% of the lines, adding noise and applying an arbitrary known transformation (rotation, scale factor, translation). The objective of this first evaluation is:

- Compare the robustness and the speed of the two optimization algorithms (simulated annealing and RANSAC) in order to choose the best to apply.
- Evaluate the performance of our algorithm using the best optimization algorithm.
- Make a comparison between the efficiency of our algorithm and the Iterative Closest line algorithm (ICL).

The various experiments carried out have proven the robustness and speed of RANSAC compared to simulated annealing as shown in Figure 5.2. The obtained results have proven the performance of our algorithm to register two sets of 3D segments, whatever the initial position, unlike ICL which required a good approximation of the initial transformation to be able to converge towards the correct solution as shown in Table 5.1 and Figure 5.3. The average running time of our algorithm for the different tests in Table 5.1 is 161.6 s by fixing the maximum number of iterations of RANSAC to 5000 iterations, knowing that the first 3D segments set contains 144 lines and the second contains 128 lines.

5.7.2 Evaluation on real data

Afterwards, we performed the registration of heterogeneous data: (terrestrial images/terrestrial LiDAR data) as shown in Figure 5.4 and (aerial images/aerial LiDAR data) as shown in Figure 5.5. As it is difficult to construct a ground truth for heterogeneous data registration, we evaluated the registration quality on the 3D visual results. The obtained results demonstrate our algorithm's ability to efficiently register image and LiDAR data. In addition, our algorithm has proven its robustness to small disturbance of verticality.

Achieving precise results requires fine-tuning of the algorithm's parameters. The distance threshold must be chosen reasonably so as not to consider a large number of lines as outliers. The number of iterations of RANSAC must be large enough to ensure the robustness of the algorithm.

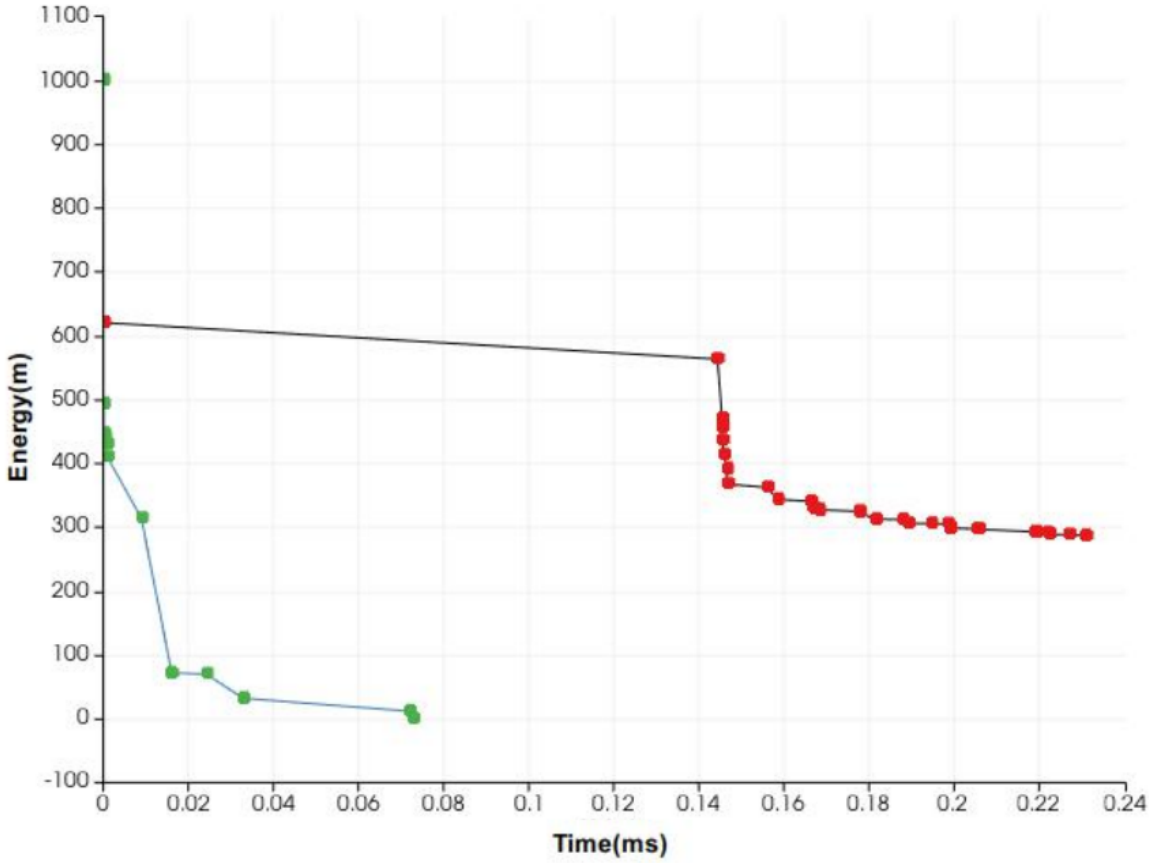
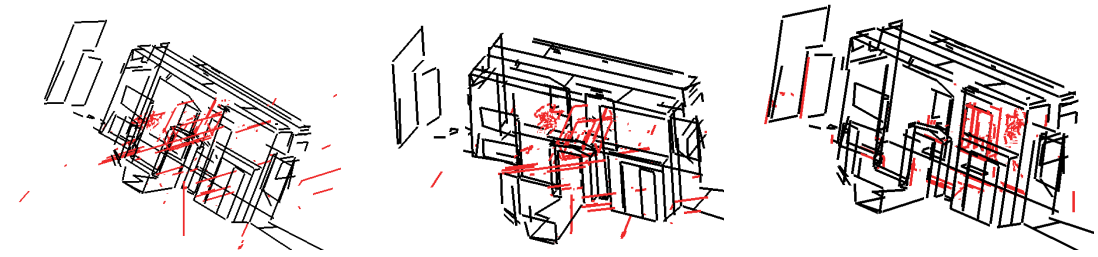


Figure 5.2: Comparison of the convergence speed and the robustness of RANSAC and simulated annealing: X axis represents time in milliseconds, Y axis represents the energy, green points represent the minimums estimated by RANSAC and the red points represent the minimums estimated by simulated annealing.

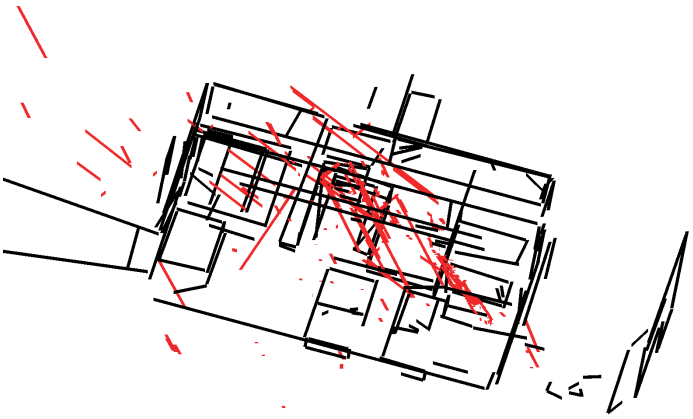
Transformation			Our result			ICL result		
Translation	Rotation	Scale (s_0)	Translation Error	Rotation Error	Scale (s)	Translation Error	Rotation Error	Scale (s)
0 m	0°	1	1.77e-15 m	1.06e-15°	$\frac{\lvert s-s_0\rvert}{s_0}$=0	3.6e-16 m	4.7e-15°	$\frac{\lvert s-s_0\rvert}{s_0}$=0
0.39 m	4.66°	0.85	0 m	0.04°	$\frac{\lvert s-s_0\rvert}{s_0}$=0	0.09 m	1.29°	$\frac{\lvert s-s_0\rvert}{s_0}$=0.005
1.31 m	15.66°	1.5	0 m	0.2°	$\frac{\lvert s-s_0\rvert}{s_0}$=0	0.20 m	15.39°	$\frac{\lvert s-s_0\rvert}{s_0}$=0.971
4.20 m	32.66°	2	0 m	0.2°	$\frac{\lvert s-s_0\rvert}{s_0}$=0	1.42 m	27.520	$\frac{\lvert s-s_0\rvert}{s_0}$=0.97

Table 5.1: Performance tests of our algorithm on synthetic data using different initial errors.

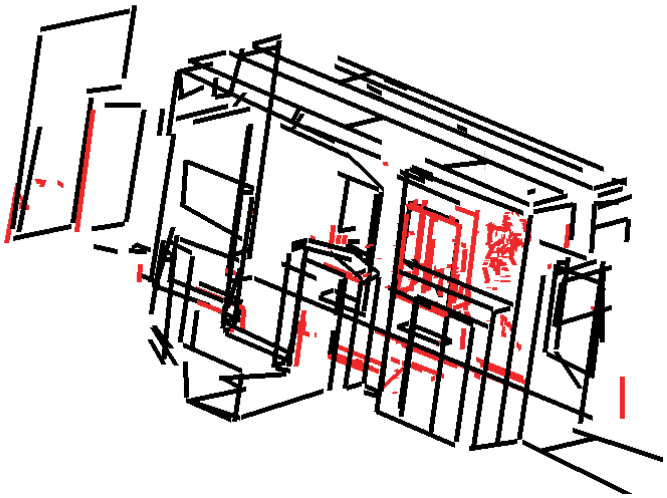


(a) (b) (c)

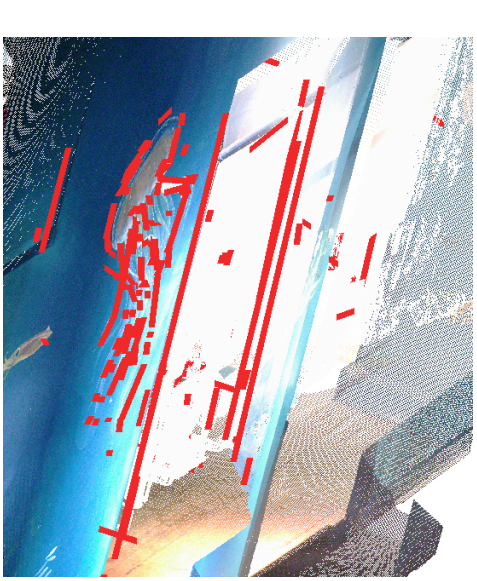
Figure 5.3: Comparison between the performance of ICL and our algorithm. (a) Initial position, (b) ICL result, (c) Our result



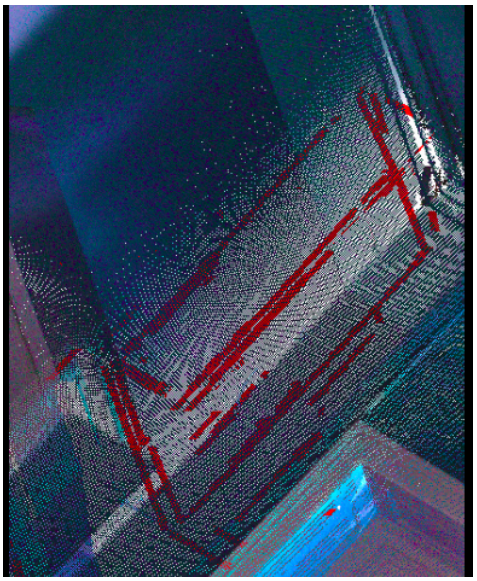
(a)



(b)



(c)



(d)

Figure 5.4: Terrestrial image/Terrestrial LiDAR registration results: (a) position of the two line clouds before registration (red:image lines, black LiDAR lines), (b) position of the two line clouds after registration, (c,d) registration of image lines and the LiDAR scan.

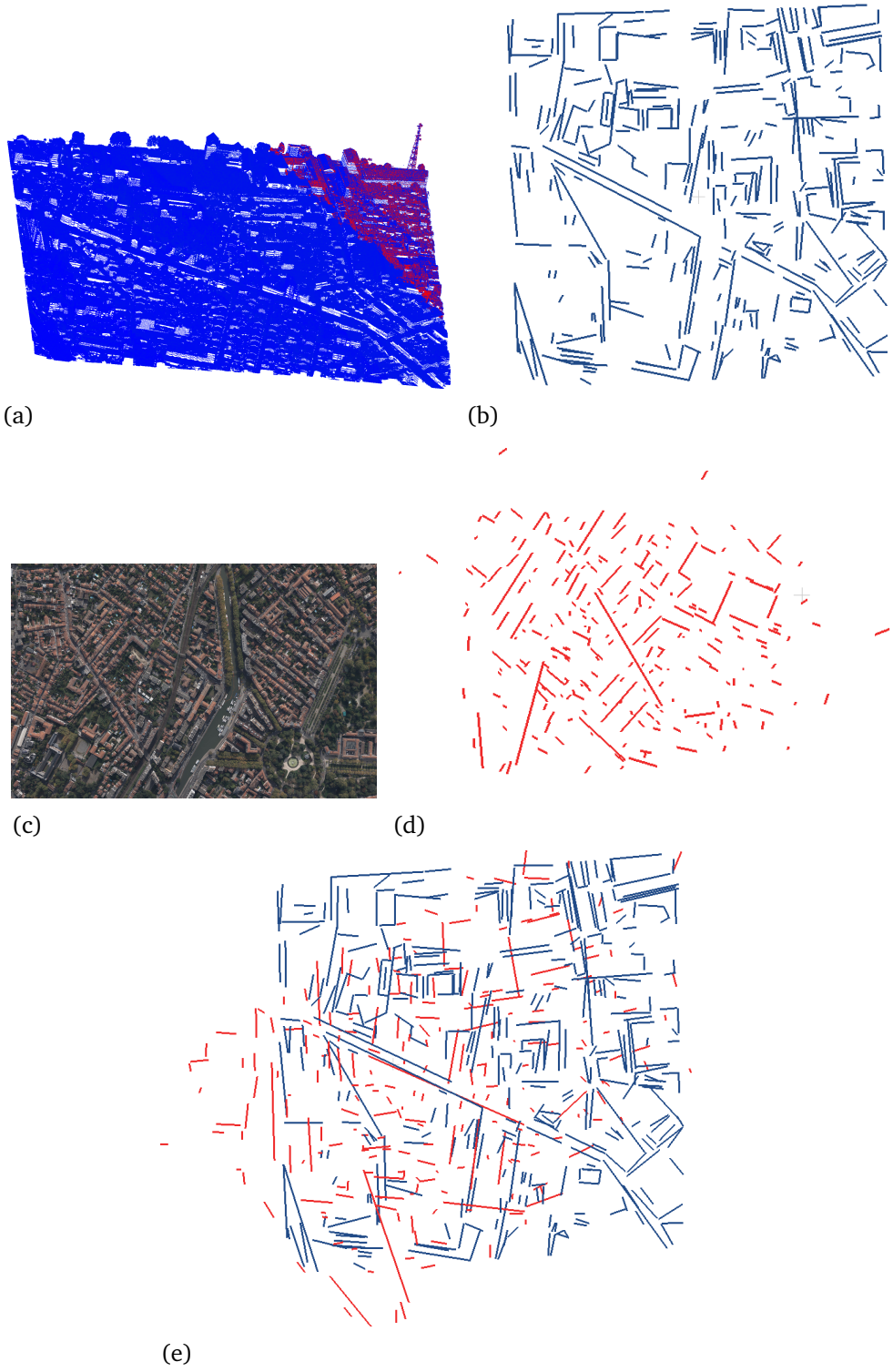


Figure 5.5: Aerial image/Aerial LiDAR registration results: (a) aerial LiDAR scan, (b) extracted lines from the aerial LiDAR scan, (c) an aerial image, (d) Reconstructed lines from an aerial image sequence , (e) Registration result.

5.8 Conclusion

In this chapter, we have dealt with the heterogeneous data (Image/LiDAR data) registration problem by proposing a new primitives based registration algorithm. Our

algorithm takes 3D segments as feature and extracts them from both LiDAR and image data with specific state of the art algorithms. We used the global robust energy between two segments sets, which has been defined in Section 4.5.1, and we proposed a robust approach to minimize this energy. In order to simplify this minimization, we started by clustering the 3D segments of each data set. The clusters are associated to find possible rotations, then 3D segments from associated clusters are matched in order to find a the translation and scale factor minimizing the defined energy. The obtained results demonstrate the efficiency and robustness of our algorithm to register heterogeneous data. Our main perspective on this work is to use planar polygons as primitives or to use combinations of more segments to have more characteristic features to match.

Chapter 6

Aerial/Terrestrial registration

6.1 Introduction

The aim of this thesis is not only to register heterogeneous data (image and LiDAR) but also to register data acquired by heterogeneous acquisition platforms (aerial/terrestrial). By using airborne platforms in an urban environment, we can see the top faces of buildings, but the side faces are still missing. Unlike Terrestrial platforms which give complete and dense information from side faces and do not give significant information from top faces. So in order to achieve a complete coverage of urban areas, the fusion of both aerial and terrestrial views is necessary. Aerial data can be acquired in the form of images or LiDAR and it is the same case for terrestrial data. So the problem of aerial/terrestrial registration can take several forms:

- Aerial image/Terrestrial image registration
- Aerial LiDAR/Terrestrial image registration
- Aerial image/Terrestrial LiDAR registration
- Aerial LiDAR/Terrestrial LiDAR registration

The developed algorithms within the framework of our thesis are able to solve the problem of aerial/terrestrial registration and deal the different possible forms of the problem as shown in Table 6.1.

	Aerial LiDAR	Aerial Image
Terrestrial LiDAR	Planar polygons based algorithm	3D segments based algorithm
Terrestrial Image	3D segments based algorithm	3D segments based algorithm

Table 6.1: Possible solutions

As we have four possible ways to solve the same problem (aerial/terrestrial registration), we have chosen to exploit one of the four (aerial image/ terrestrial image registration) for the following reasons:

- we have already studied the LiDAR/LiDAR registration in the first part of our thesis (interior/exterior registration).
- we have already studied the image/LiDAR registration in the second part of our thesis.
- As the 3D segments based algorithm has proven its effectiveness in solving the problem of image/LiDAR registration and LiDAR/LiDAR registration(openings based registration), we hope to test it in solving the image/image registration problem.

6.2 State of the art

As the features of aerial and terrestrial data are complementary, several research works are focused on the exploration of joint data processing methods to overcome the shortcomings of methods based on aerial or terrestrial data alone. Aerial and terrestrial data can be acquired as images or 3D point clouds. Integrating data from these two acquisition platforms requires a registration step. Several works in the literature have addressed the problem of aerial/terrestrial registration.

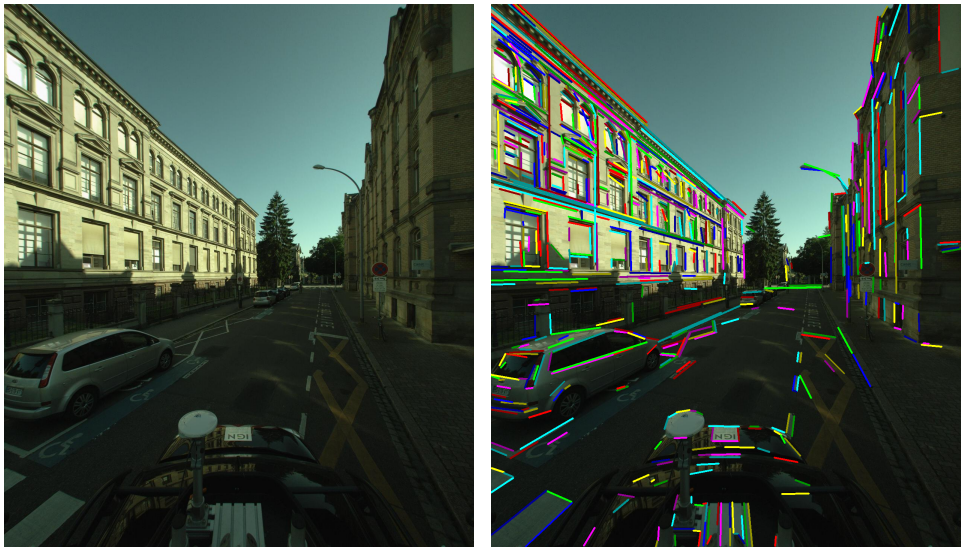
An efficient method for the registration of airborne and terrestrial mobile LiDAR has been proposed in (Teo and Huang, 2014). It used the least squares 3-D surface matching algorithm to minimize the Euclidean surface distance between two datasets. An algorithm based on line features has been proposed in (Von Hansen et al., 2008) which represents a potential solution for terrestrial and airborne LIDAR data registration. The basic idea of this algorithm consists in using orientation histograms to find the rotation and a generate and test scheme to find the translation

parameters by matching all possible combinations of two line segments. The authors of (Cheng, Cheng, Li and Ma, 2018) have been proposed a new solution that can achieve automatic horizontal registration with Airborne Laser Scanning (ALS) and TLS data based on building contour features extracted from both data. A coarse registration has been carried out using the four-point congruent method. Then a refinement of the horizontal registration of the building outlines is conducted. A hierarchical method for registration of ALS and terrestrial image-based point clouds has been proposed in (Baghani et al., 2018). In order to overcome the problems of the line-based methods, a novel polar parameterized mathematical model was extended for horizontal matching. An innovative target composed of three perpendicular planes that combine the properties of plane and plane and volume target has been implemented in (Urbančič et al., 2019). This new target enables the precise determination of reference target points in aerial and terrestrial point clouds. This property make it suitable for accurately registering point clouds produced from Unmanned Aerial Vehicle (UAV) images and terrestrial laser scans.

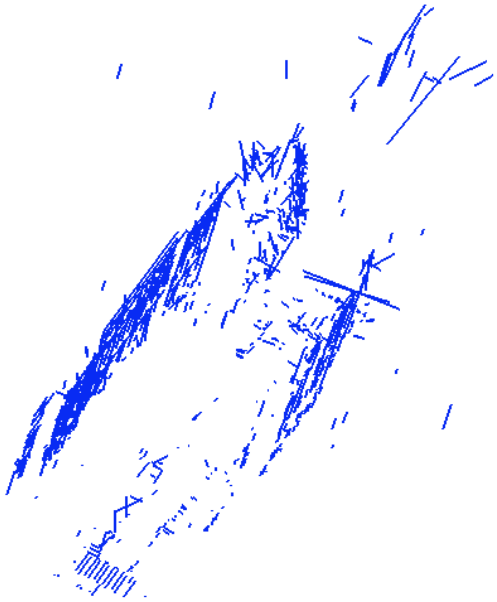
6.3 3D segment extraction from heterogeneous image data

6.3.1 3D segment extraction from terrestrial image sequence

In order to extract 3D line segments from terrestrial image sequence, we chose to work with the method proposed in Section 3.6.3. This method uses an oriented image sequence as input, whose camera poses can be obtained by any conventional Structure From Motion (SfM) pipeline. It generates a final line-based 3D model by clustering 2D segments from different views. These 2D segments are detected using our improved version of MLSD as shown in Figure 6.1.



(a) (b)

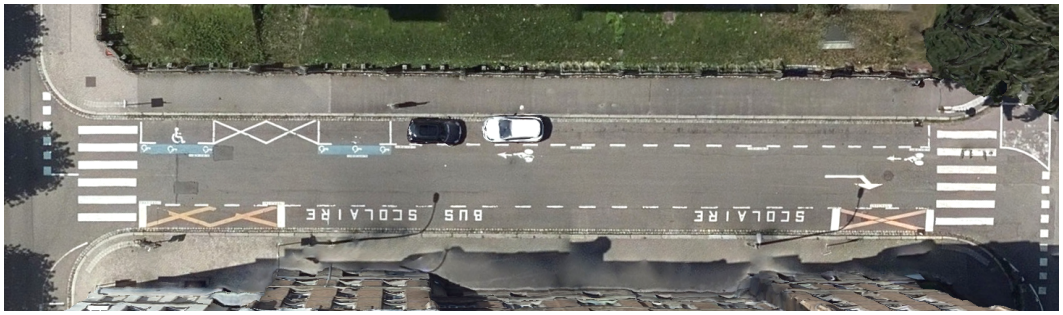


(c)

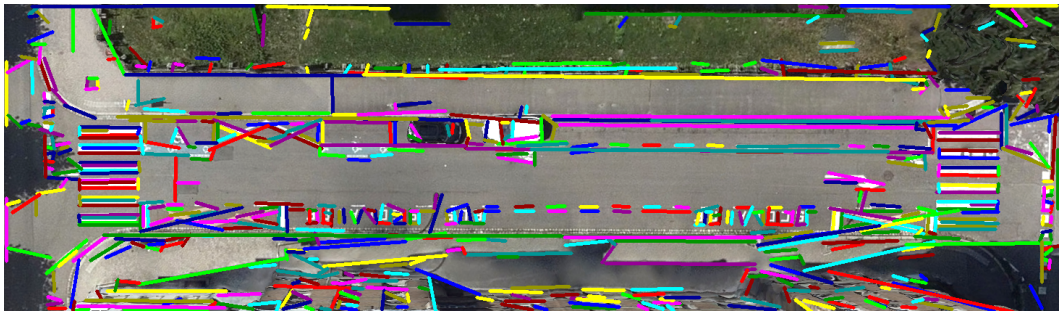
Figure 6.1: Reconstructed 3D line cloud from a Terrestrial image sequence: (a) An image from the terrestrial image sequence,(b) MLSD detected lines, (c) The reconstructed lines.

6.3.2 3D Line segments detection and reconstruction for an orthoimage

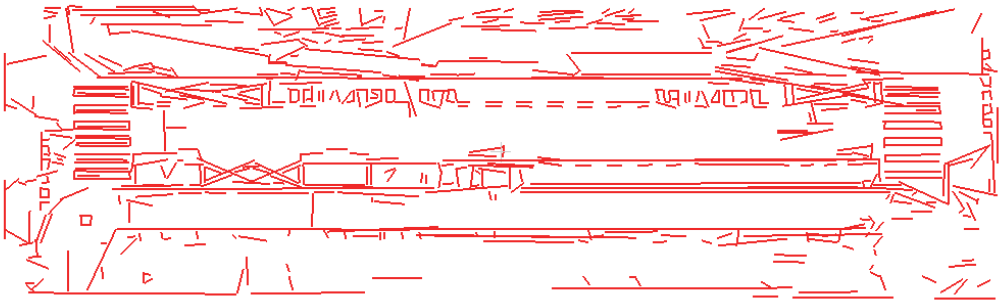
For aerial data, we have a single orthoimage. In order to get 3D line segments from this image, we started by detecting 2D lines using our improved version of MLSD. Afterwards, we got a 3D model from each 2D line by considering that the z coordinate equal to 0 as shown in Figure 6.2



(a)



(b)



(c)

Figure 6.2: Reconstructed 3D line cloud from an orthoimage: (a) The orthoimage,(b) MLSD detected lines, (c) The reconstructed lines.

6.4 3D segments based algorithm adaptation

The 3D segments based algorithm has proven its efficiency to solve the image/LiDAR registration problem, it can be a potential solution for aerial image/terrestrial image registration. This algorithm takes two 3D segments sets as input. It starts by clustering the 3D segments of each data-set according to their directions. Afterwards, it associates the obtained clusters to find possible rotations. Then 3D segments from associated clusters are matched in order to find the translation and scale factor minimizing a global robust energy between the two segment sets. We have applied the following adaptations on this algorithm to make it able to solve the current registra-

tion problem.

6.4.1 Valid cluster associations

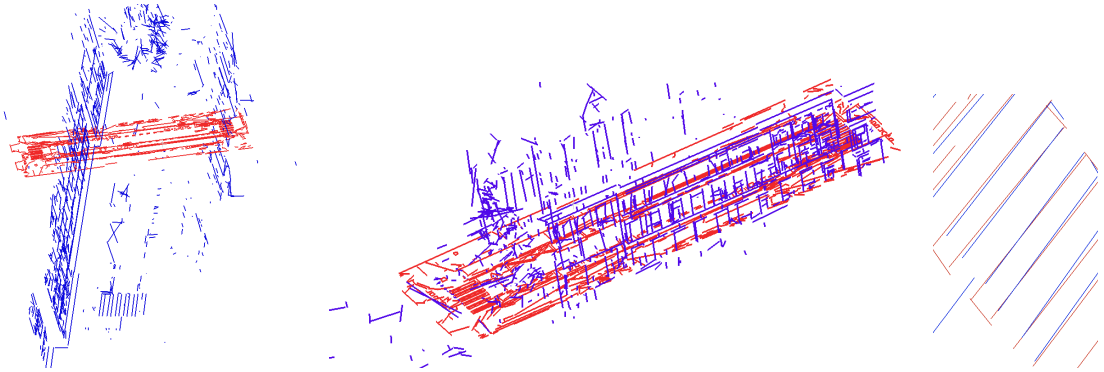
To define a rotation between the two segment sets, it is sufficient to associate two pairs of segments from the two sets. However this is not very precise, so we propose to associate the segment clusters. As for aerial data most detected lines lie on the (horizontal) ground, we do not have a vertical cluster, so we have removed the vertical cluster from terrestrial data. Now, we can associate any pair of aerial clusters to any pair of terrestrial clusters if they have a compatible angle.

6.4.2 RANSAC optimization

For the original algorithm, at each RANSAC iteration, we randomly select a valid cluster association, then we randomly select one 3D segment in each of the associated clusters. For the aerial data, we have chosen to work with an orthoimage, where all the segments are coplanar. In this situation, if we only use two segments pairs, we obtained a degenerate scale estimation. By adding a third segment pair, we can get additional information about the distance between segments which can help to correctly estimate the scale factor.

6.5 Evaluation and discussion

As it is difficult to construct a ground truth for aerial/terrestrial data registration, we evaluated the registration quality on the 3D visual results. By applying our algorithm, we could register an aerial segment set composed of 271 lines and a terrestrial segment set composed of 846 lines. By fixing the iteration number to 2000 iterations, we got the final registration result after 762.459s. The obtained results demonstrate our algorithm's ability to efficiently register aerial and terrestrial data as shown in Figure 6.3. By inspecting the pedestrian path, we can clearly see that the registration error does not exceed a few centimeters



(a) (b)

Figure 6.3: Aerial/Terrestrial registration: 3D lines reconstructed aerial images(red), 3D lines reconstructed from terrestrial images (blue)
(a): the initial position, (b): positions after registration, (c): position of pedestrian path lines after registration

6.6 Conclusion

In this chapter, we have addressed the problem of aerial/terrestrial registration that can take several forms depending on the modality (image or LiDAR). We have chosen to address the problem of registration of aerial image / terrestrial image. The very strong difference in point of view makes it difficult to extract comparable features between the two data. Conversely, with our 3D segments and our RANSAC-based approach, we assume that the best transformation is the one which minimizes the global robust energy between all the extracted 3D segments and therefore which is the most consistent with all the information of the data-sets to be registered. The adaptation of our 3D segments based algorithm has made it very effective in dealing with this difficult problem.

Chapter 7

Conclusion

7.1 Contributions

The aim of this PhD was to address all registration issues faced by the BIOM project, as a major objective of the project is exploiting jointly data from various sensors (images/LiDAR) and viewpoints (terrestrial indoor or outdoor, aerial). The work carried out has confirmed that the environment and the type of data drive the choice of the registration algorithm. So, our objective was to explore fundamental properties of the data and the acquisition platforms in order to propose potential solutions for all the registration problems encountered by the project. As in a building environment, most objects are composed of geometric primitives (planar polygons, straight lines, openings), we chose to introduce registration algorithms based on these primitives. The basic idea of these algorithms consists in the definition of a global robust energy between the extracted primitives from the data sets to register and the proposal of a robust method for optimizing this energy based on the RANSAC paradigm. Our solutions have exceeded the limitations of existing algorithms and have proven their effectiveness in solving the challenging problems encountered by the project. The registration of indoor and outdoor scans represents a major challenge for indoor and outdoor building modeling. The lack of overlap between indoor and outdoor data is the most challenging obstacle, more so when both data sets are acquired separately and using different types of sensors. In order to solve this problem, we started by

proposing two solutions separately:

- A first solution based on polygon detection and matching. The strong points of this algorithm lies in the fact that it exploits the very small overlap between indoor and outdoor scans of the same building due to the property of laser rays crossing detected façades.
- A second solution based on openings, which represent the unique common entities that can be seen from inside and outside. Therefore, it can help the registration of indoor and outdoor point clouds.

Given the obtained results, we can consider that the two proposed algorithms are potential solutions for the indoor/outdoor registration problem, though they have some limitations. In order to overcome these limitations, we have combined the two proposed algorithms. This combination produced an efficient hybrid algorithm able to perform indoor/outdoor registration with great precision.

Image/LiDAR registration and aerial/terrestrial registration are very hard problem. The strong difference in point of view and modality makes it difficult to extract comparable features between the two modalities. Conversely, with our primitives (3D segments) and our RANSAC-based approach, we assume that the best transformation is the one which minimizes the global robust energy between all the extracted primitives and therefore which is the most consistent with all the information of the data sets to register. The obtained results have proven the ability of our 3D segments based algorithm to deal with this kind of registration problems. Table 7.1 represents a synthesis of our contributions.

Problem		Solution
Primitives detection	Planar polygons detection	Adaptation of two algorithms (RANSAC based sensor topology, MSAC)
	Openings detection	Proposal of an efficient algorithm
	3D line segments Detection (LiDAR data)	Selection and use of an existing algorithm (3DLineDetection)
	3D line segments reconstruction (image sequence)	Improvement of an existing algorithm (Line3D ++)
Registration	Indoor/outdoor registration	Proposal of a planar polygons based algorithm
		Proposal of an openings based algorithm
		Proposal of a hybrid solution (planar polygons + openings)
	Image/LiDAR data registration	Proposal of a 3D segments based algorithm
	Aerial/ Terrestrial registration	Adaptation of our 3D segments based algorithm

Table 7.1: Synthesis of our contributions

7.2 Future work

The studied problems in this thesis, the proposed solutions and the obtained results open several perspectives:

- **Openings detection:**

  We can allow more flexible shapes for openings, for instance by extracting corner points for each connected component and connecting them using parametric B-spline curves. We can also perform a semantic segmentation of the input point clouds to improve the accuracy and robustness of openings detection.

- **Indoor/outdoor registration:**

  Our main perspective concerning this part is to extract outdoor points seen from the indoor scans in order to get additional information, if some are present.

- **Image/LiDAR registration:**

  Our main perspective on this part is to use combinations of more segments to

have more characteristic features to match.

- **Aerial/Terrestrial registration:**

  Our main perspective for this difficult task is to exploit the different proposed algorithms to deal with the different modalities that can take the problem. For instance, we can perform the aerial LiDAR/terrestrial LiDAR registration using planar polygons as features. We can also solve the aerial LiDAR/terrestrial image registration or the aerial image/terrestrial LiDAR registration problem using 3D segments as features.

Appendices

Appendix A

Implementation

All the developed algorithms in this work have been implemented in C++ using the following libraries:

- CGAL library `https://www.cgal.org/`: to carry out the geometric calculations.
- Boost library `https://www.boost.org/`: to calculate polygons intersection.
- Eigen library `https://eigen.tuxfamily.org/dox/` for matrix calculation.
- PCL library `https://pointclouds.org/` for point cloud processing.
- OpenCV library `https://opencv.org/` for image processing

Appendix B

MLSD Improvement

B.1 Introduction

The traditional line segment detection algorithms (Canny edge, Hough transform) are slow and generate many false detections. To correct and overcome these problems, a new family of algorithms based on a contrario theory has been proposed. These algorithms have become the current state of art methods for line detection. Using a contrario theory, the algorithms can automatically define whether a line is meaningful with a score called Number of False Alarms. The first proposed algorithms in this category are EDLine (Akinlar and Topal, 2011) and LSD (Von Gioi et al., 2012). Later, a multi-scale extension of LSD (MLSD) was proposed in (Salaün et al., 2016). This appendix is devoted to the explanation of the general principle of LSD and MLSD as well as the illustration of our improvements applied to the two detectors in order to boost their performance. Before that, we introduce and explain the general principle of the a contrario theory.

B.2 A contrario method

According to (Desolneux, 2016), the a contrario method is a statistical approach based on hypothesis testing. This approach is used in order to detect geometric meaningful events in images. The basic idea consists in computing the probabil-

ity of an observed geometric event under a noise model (null hypothesis) H_0 (also called, in some papers: the a contrario noise model, the background model or the naive model) and then declare the event meaningful when this probability is small enough. Generally, the independent uniform distribution over the considered elements is represented by the noise model.
The general principle of this approach can be summarized as follows:

- Define a null hypothesis H_0 that is a probability distribution on elements, and take a small number ϵ.
- Observe a geometric event (that is a configuration of elements) in an image, and denote it by E.
- Compute the probability of E under H_0.
- If this probability is smaller than a threshold computed to ensure that on the average there are less than ϵ detections in an image where the elements would be distributed according to H_0, then declare E as ϵ-meaningful.

B.3 Line Segment Detector (LSD)

LSD is considered one of the best and most popular methods among proposed line detectors. It accurately detects segments and does not use any threshold tuning, relying instead on the a contrario methodology.

Following (Von Gioi et al., 2012), LSD starts by extracting connected regions, called line support regions. This extraction is based on level-line angle computed at each pixel, in order to produce a level-line field, and the segmentation of this field into connected regions of pixels that share a similar level-line angle up to a certain tolerance T, as shown in Figure B.1. Each line support region is considered as a candidate for a line segment. The corresponding geometrical object (a rectangle in this case) must be associated with it. The main rectangle direction will be presented by the principal inertial axis of the line support region. In the rectangle, the pixels whose level-line angle corresponds to the angle of the rectangle, up to a tolerance T, are called aligned points see Figure B.2 Each rectangle must go through a validation pro-

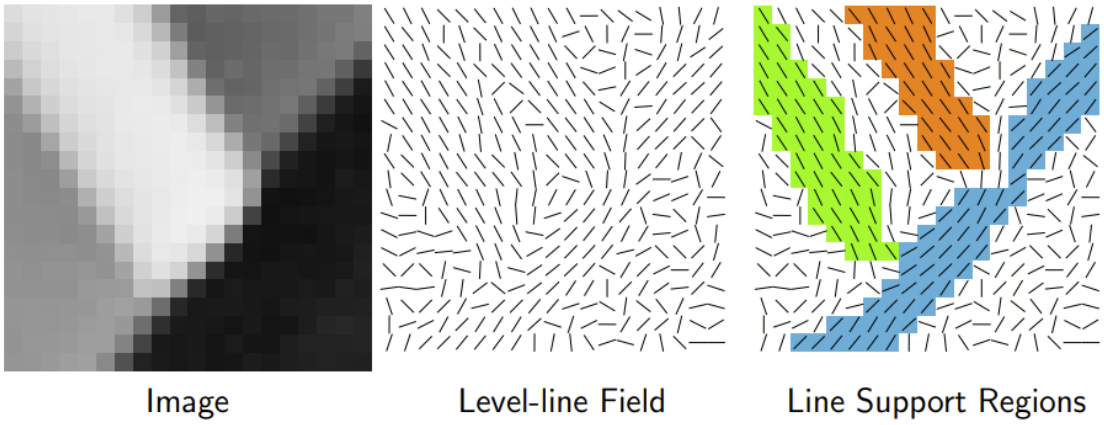


Figure B.1: Line Support Regions illustration (Von Gioi et al., 2012)

cedure after computing the total number of pixels in the rectangle n and its number of aligned points k.

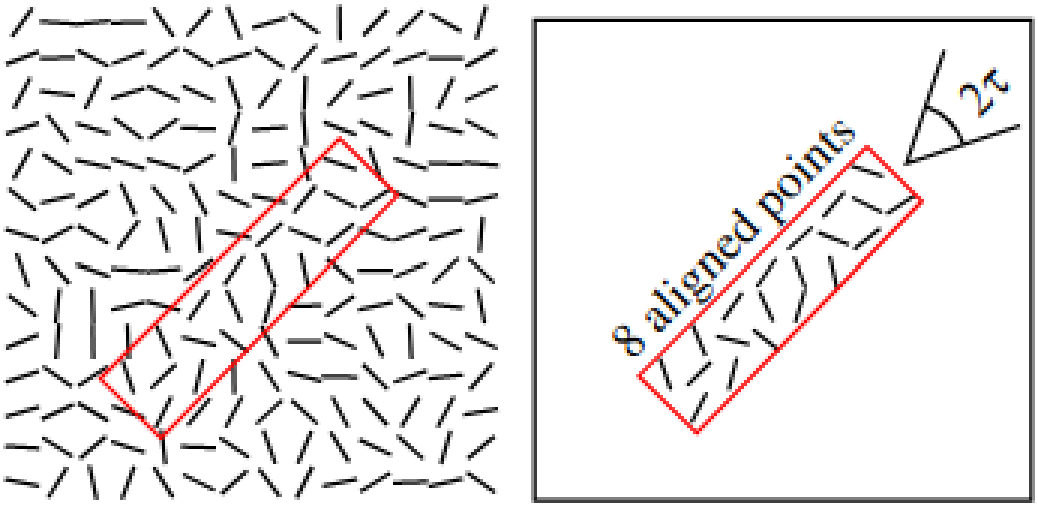


Figure B.2: Aligned points illustration (Von Gioi et al., 2012)

The validation step is based on the a contrario approach and it takes into consideration the number of aligned points to inform the decision. Given an image i and a rectangle r, we will note $k(r, i)$ the number of aligned points and $n(r)$ the total number of pixels in r. Then, the expected number of events which are as good as the observed one is

$$N_{Test} \cdot P_{H_0}[k(r, I) \geq k(r, i)], \tag{B.1}$$

where the number of tests N_{Test} is the total number of possible rectangles being considered, P_{H_0} is the probability on the a contrario model H_0 and I is a random image following H_0. Under hypothesis H_0, the probability that a pixel on the a contrario

model is an aligned point is:

$$p = T/\pi \tag{B.2}$$

As $k(r, I)$ follows a binomial distribution, we can write the probability term as the following:

$$P_{H_0}[k(r, I) \geq k(r, i)] = B(n(r), k(r, i), p) \tag{B.3}$$

where $B(n, k, p)$ is the tail of the binomial distribution:

$$B(n, k, p) = \sum_{j=k}^{n} \binom{n}{j} p^i (1-p)^{n-j}. \tag{B.4}$$

The exhaustive choice is to take all the rectangles starting and ending at image pixels. In an $N \times M$ image this gives $NM \times NM$ different rectangles. Also, $\sqrt{NM}$ different width values are considered for each one as shown in Figure B.3 Finally, the Number

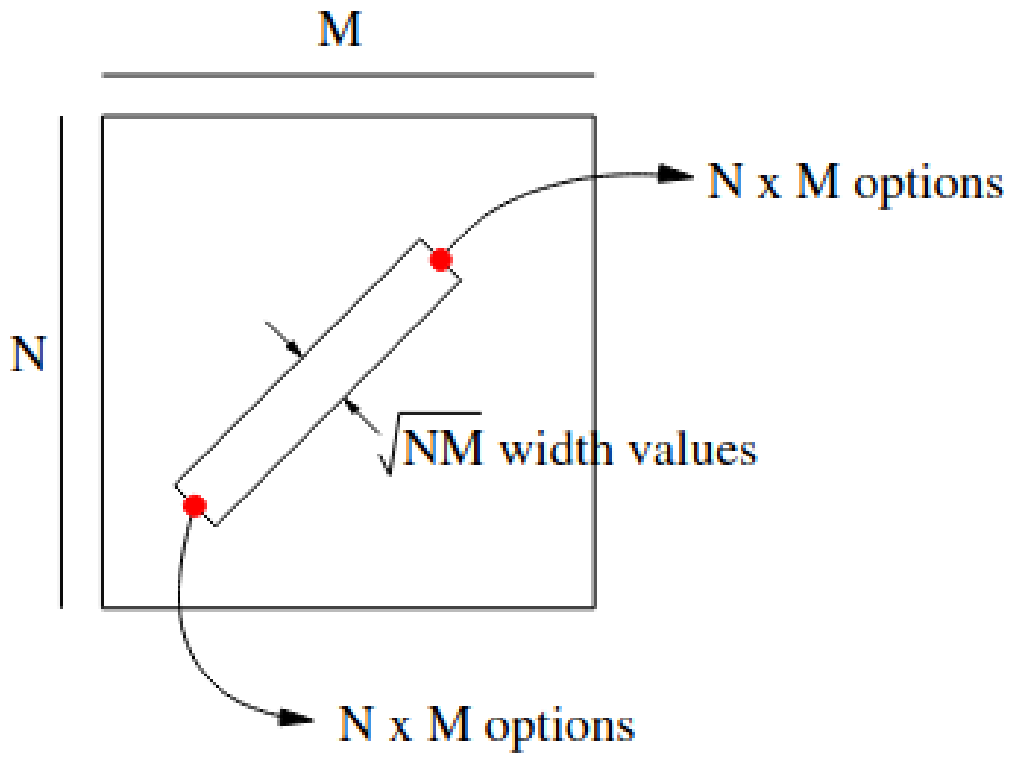


Figure B.3: Estimation of the number of tests. (Von Gioi et al., 2012)

of False Alarms (NFA) associated with a rectangle r on the image i is defined as:

$$NFA(r, i) = (NM)^{\frac{5}{2}} \lambda B(n(r), k(r, i), p) \tag{B.5}$$

where, $(NM)^{\frac{5}{2}}$ is the number of rectangles, λ is the number different p values poten-

tially tried (The precision p is initially set to the value $\frac{T}{\pi}$ but other multiples are also tested to cover the relevant range of value).

B.4 LSD parameters improvement

B.4.1 Reduce Region Radius

In some cases, the T angle-tolerance method produces a wrong interpretation. This problem can arise when two straight edges are present in the image forming an angle between them smaller than the tolerance T as shown in Figure B.4 The proposed

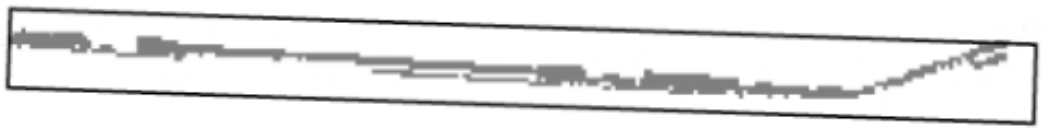


Figure B.4: An example of region growing problems

solution of this problem in LSD is to remove gradually the pixels that are farther from the seed point until the criterion is satisfied or the region is too small and rejected. At each iteration, the algorithm removes the farthest pixels of the region to reduce the region's radius to 75 % of its value.

Improvement

Reducing the radius to 75 % from its previous value at each iteration will penalize some pixels that are good candidates. We propose to reduce the radius to 95 % of its value instead 75 %. This solution will:

- Solve the problem that can arise at the region growing process.
- Allow a large number of pixels to be examined.

Convolution with a Gaussian kernel

The scaling is performed by a Gaussian sub-sampling: the image is filtered with a Gaussian kernel to avoid aliasing and then sub-sampled. The standard deviation of

the Gaussian kernel is fixed to 0.8 and the size to 7×7. As we know, the size of the kernel used to blur an image can have a significant effect on the result of the blurring and any downstream analysis carried out on the blurred image. In addition, larger σ values may remove more noise, but they will also remove details from an image. A σ equal to 0.8 and a kernel size equal to 7×7 is not a good choice. So for better effect, we set σ to 1.6 to be able to remove more noise and preserve image details. This choice of the σ value leads us to fix the kernel size at 11×11, because the kernel radius must be equal to $3\sigma + 1$.

B.5 Multi-scale line segment detector (MLSD)

(Salaün et al., 2016) propose a multi-scale extension of LSD. This multi-scale nature makes it much less prone to over-segmentation, more robust to low contrast and less sensitive to noise, while keeping the parameter-less advantage of LSD and still being fast.

MLSD uses the same background model as generally used in a contrario methods. The authors of MLSD have seen that the size of the picture matters a lot for segment detection. A smaller version of the same picture often yields fewer but proportionally longer segment detections, while the original picture may yield a lot of fragmented segments. They thus use a multi-scale approach to find long segments at coarse scales, refining their location at a finer scale. At each scale, They consider new segment candidates at the same locations than segments at the previous scale, possibly merge them using the multi-segment criterion, and keep the resulting segments that pass the NFA condition. At the coarsest scale, they only detect segments classically with LSD. At subsequent scales, they first use information from the previous scale to find new segments.

B.6 MLSD improvements

B.6.1 Angle tolerance

For empirical reasons, MLSD has set the value of angle tolerance to $\frac{\pi}{8}$, as for LSD. LSD uses an 8-connected neighborhood for region growing. The tolerance T is set to $\frac{\pi}{8}$ radian. It is a reasonable choice which makes the probability that a pixel is aligned point at $\frac{1}{8}$.

For MLSD, at the coarsest scale, we only detect segments classically with LSD. At subsequent scales, we first use information from the previous scale to find new segments. So, the components can belong to the same line, or to parallel and close lines that were merged at a coarser scale. In this case a tolerance angle of $\frac{\pi}{8}$ is too fine and will eliminate some good segments that can be merged, this is why we have observed that taking a tolerance angle of $\frac{\pi}{4}$ remains a reasonable choice and allows to test a very large set of merged segments.

Convolution with a Gaussian kernel

For the pyramid image construction, MLSD smooths each scaled image with a Gaussian kernel. It fixes the σ value to 0.8 and kernel size to 3×3. At each scale i, MLSD multiplies the initial σ value by 2^i. So, for some scales we have a small sigma value equal to 0.8, which can not eliminate more noise. For other scales, σ value is very big, which can eliminate more noise but remove more details from the image. To solve this problem, we proposed to fix the σ value for all the scale to 1.6. This value is not small enough (to be able to remove noise) and is not large enough (to not remove details in the image). This choice of the σ value leads us to fix the kernel size at 11×11, because the kernel radius must be equal to $3\sigma + 1$.

Bound to quantization error

According to (Kim et al., 2017), interpolation while constructing the image pyramid usually causes a difference between the gradients of the original image and the scaled image. Pixels with small gradient magnitude correspond to flat zones. Also, they naturally present a higher error in the gradient direction computation due to

the quantization of their values. In LSD and MLSD the pixels with gradient magnitude smaller than a threshold D are therefore rejected. LSD and MLSD compute the threshold D as

$$D = \frac{q}{sin(T)}, \tag{B.6}$$

where T is the tolerance angle and q is a bound on the possible error in the gradient value due to quantization effects. For empirical reasons, LSD and MLSD fix the q value to 2. From our point of view, using a fixed value for all the scales is not a good solution due to gradient variation. Our proposed solution consists in using different values for this parameter at each scale. For each scaled image, the value of this parameter can be defined according to the noise level.

B.7 Evaluation and discussion

For the moment the evaluation of our improved versions of MLSD and LSD focuses on the visual aspect, i.e. is a qualitative evaluation. This evaluation was carried out on images of different sizes.

B.7.1 Big image

The first evaluation was performed on a big image of 18 Mpixels (size 5184×3456) as shown in Figure B.5.

LSD vs improved LSD

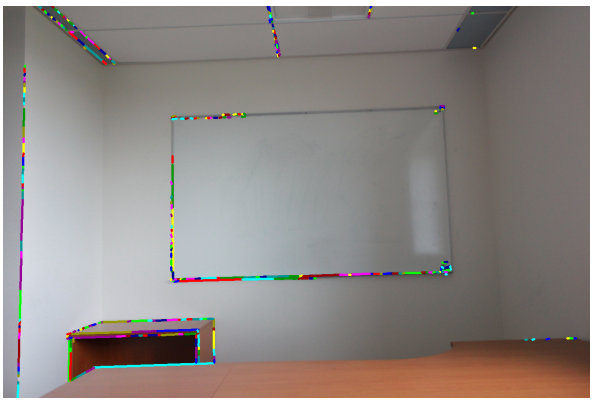
With the two versions of LSD, we could not detect all the possible lines. The detected lines with LSD are very short as shown in Figure B.5(a). Some detected lines have become longer with the improved version as shown in Figure B.5(b).

MLSD vs improved MLSD

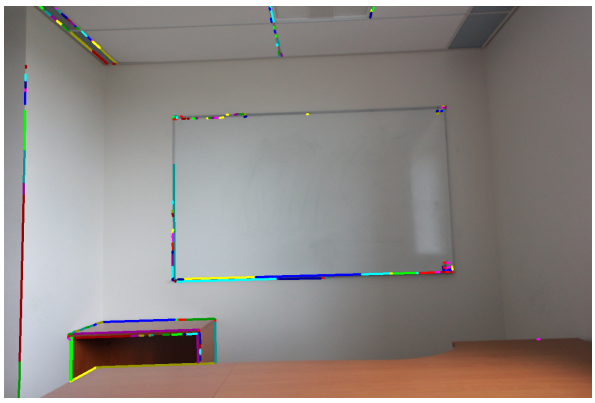
With the two versions of MLSD, we have detected more lines than LSD. The detected lines with MLSD are longer than the detected lines by lsd by they are interrupted as shown in Figure B.5(c). With the improved MLSD, we have detected more lines than MLSD and most of them are complete as shown in Figure B.5(d).

Improved LSD vs improved MLSD

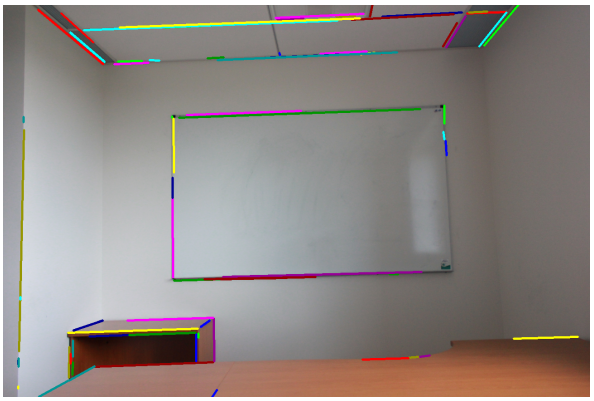
By comparing the two figures B.5(b) and B.5(d), we can well see that the improved version of MLSD gives a better result compared to the improved version of LSD which leads us to deduce that multiscale detection is preferred for large images.



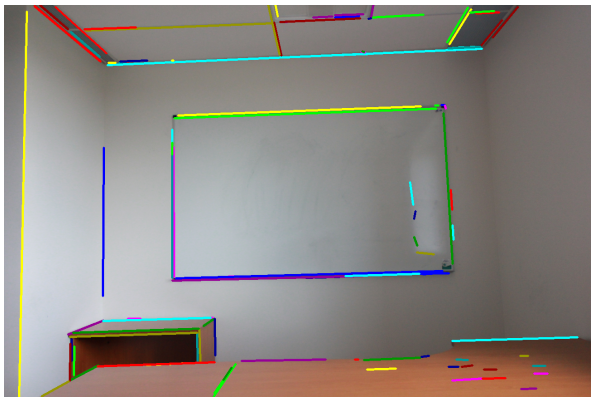
((a)) LSD



((b)) Improved-LSD



((c)) MLSD



((d)) Improved-MLSD

Figure B.5: Comparison between the performance of LSD and MLSD before and after improvement on a big size image

B.7.2 Small size images

The second evaluation was performed on a small image of size 570×461 as shown in Figure B.6.

LSD vs improved LSD

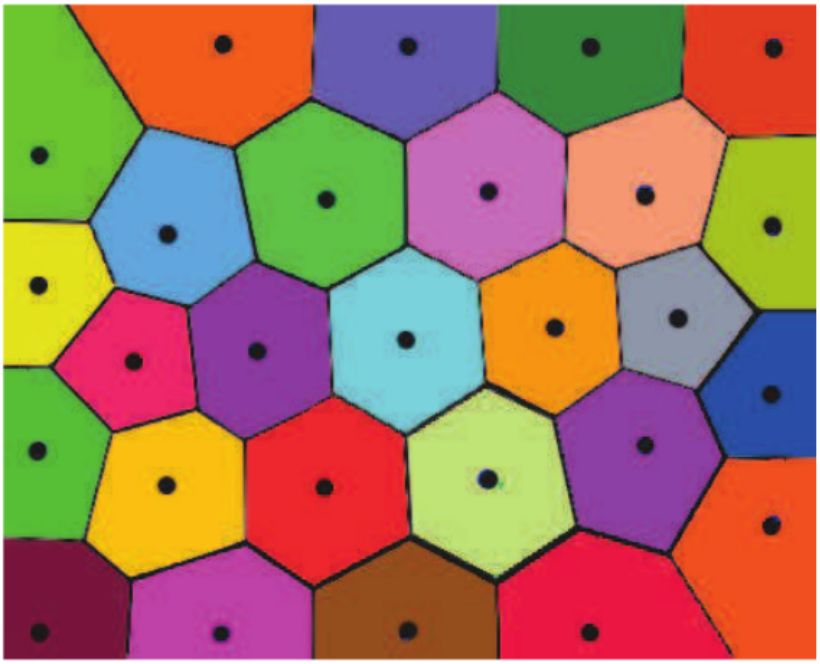
Both versions of LSD are able to detect almost all possible lines. Some detected lines with LSD are interrupted as shown in Figure B.6(b). For the improved LSD, most detected lines are complete as shown in Figure B.6(c). Neither version could detect all border lines.

MLSD vs improved MLSD

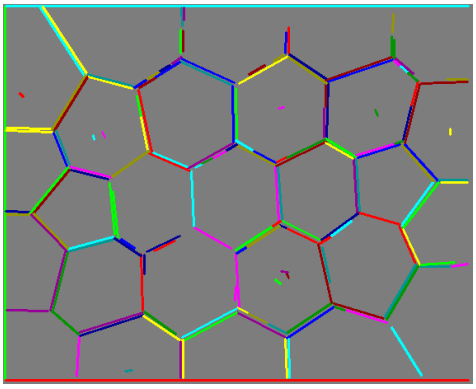
MLSD could not detect all possible lines and most detected lines are interrupted as shown in Figure B.6(d). The improved MLSD is able to detect all possible lines and the vast majority are complete as shown in Figure B.6(e).

Improved LSD vs improved MLSD

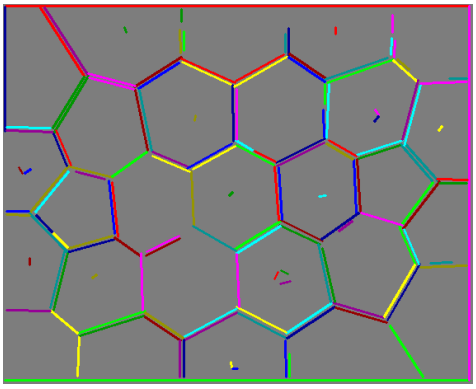
By comparing the two figures B.6(c) and B.6(e), We can well see that the improved LSD and the improved MLSD give almost the same results with the exception that MLSD could detect all the edges of the image.



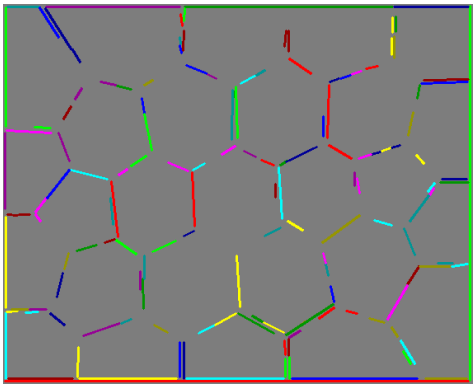
((a)) Tested image



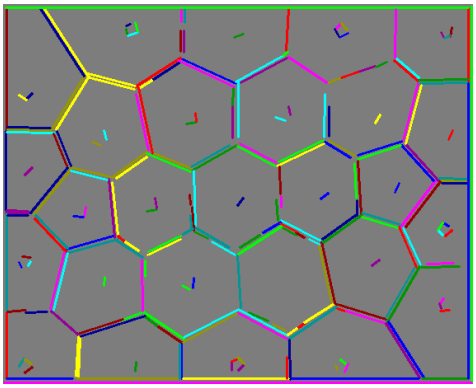
((b)) LSD



((c)) Improved-LSD



((d)) MLSD



((e)) Improved MLSD

Figure B.6: Comparison between the performance of LSD and MLSD before and after improvement on a small size image

B.7.3 Medium size images

The third evaluation was performed on two medium images: the first of size 1253×1449 as shown in Figure B.7 and the second of size 1600×940 as shown in Figure B.8.

LSD vs improved LSD

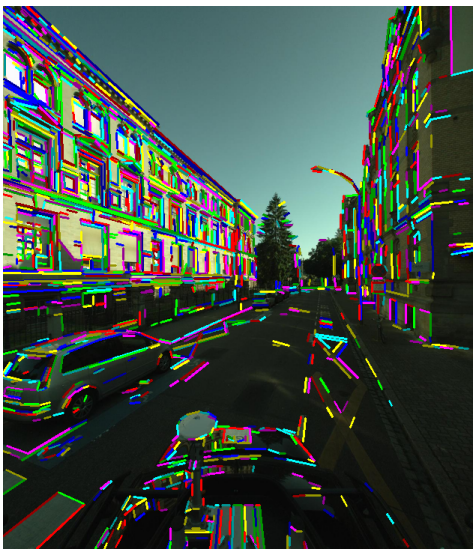
LSD has detected short and interrupted lines as shown in Figures B.7(a) and B.8(a). With the improved LSD, some new longer lines are detected as shown in Figures B.7(b) and B.8(b).

MLSD vs improved MLSD

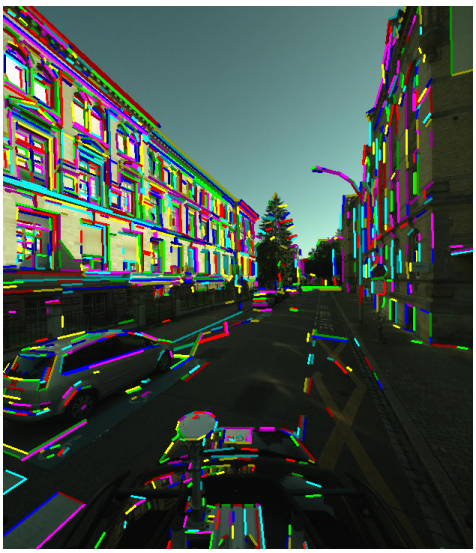
By applying MLSD, we get some parasite lines as shown in Figure B.7(c) and several missing lines as shown in the left part of Figure B.8(c). By applying the improved MLSD, we get a better result (long lines detected correctly) as shown in Figures B.7(d) and B.8(d).

Improved LSD vs improved MLSD

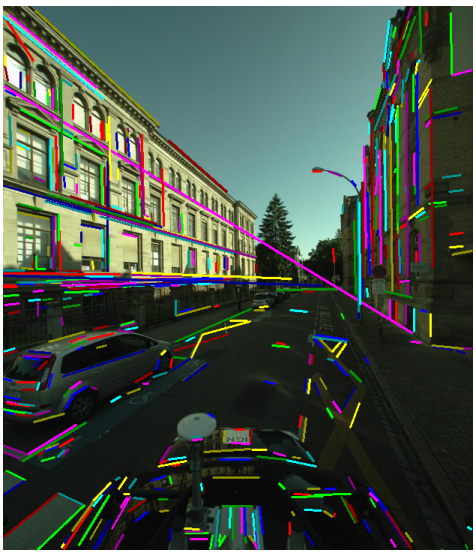
By comparing the four figures B.7(b), B.7(d), B.8(b) and B.8(d), we can well see that the improved LSD has detected short and interrupted lines unlike MLSD which has detected several long and complete lines.



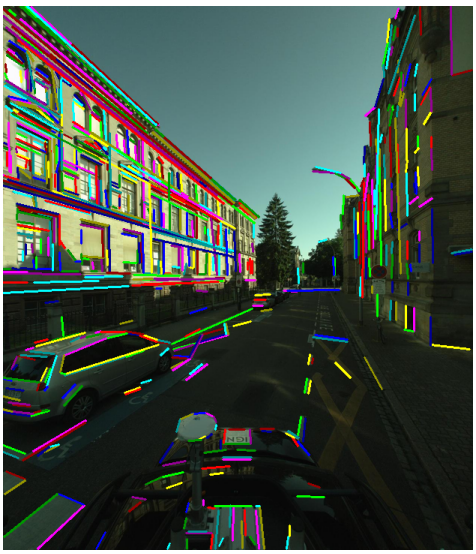
((a)) LSD



((b)) Improved LSD

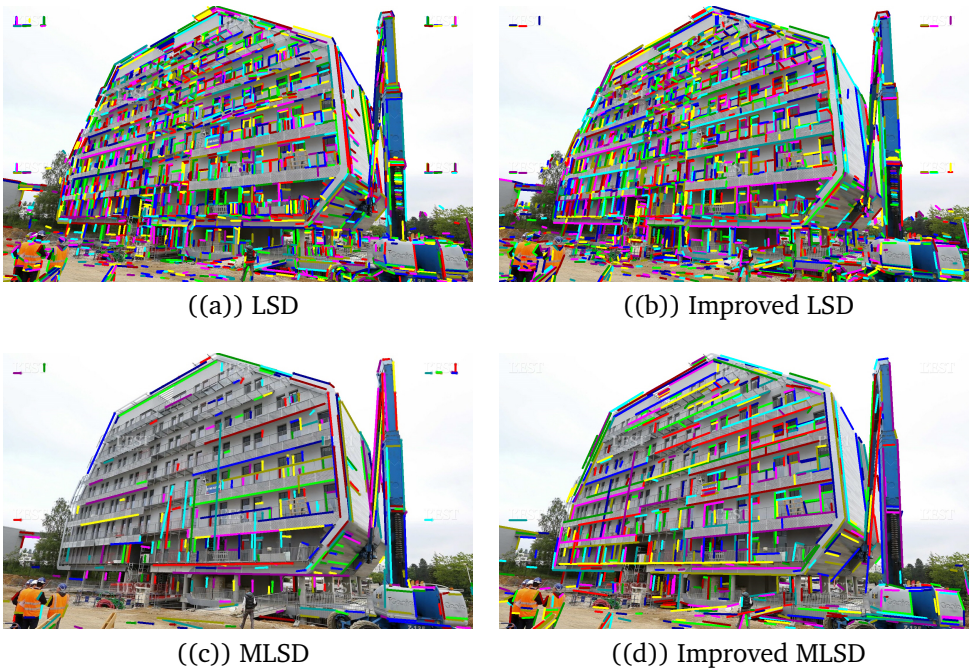


((c)) MLSD



((d)) Improved MLSD

Figure B.7: Comparison between the performance of LSD and MLSD before and after improvement on a medium size image



((a)) LSD ((b)) Improved LSD

((c)) MLSD ((d)) Improved MLSD

Figure B.8: Another comparison between the performance of LSD and MLSD before and after improvement on a medium size image

B.8 Conclusion and future works

In this appendix, we have proposed parametric improvements for two existing algorithms LSD and MLSD. the obtained results showed that the improved versions of the two algorithms exceeded the limits of the original versions. For the medium and big images, the improved MLSD represents a potential solution to detect 2D segments. For the small images, The improved LSD has also proven its effectiveness but improved MLSD remains at the top of the tested detectors. We propose as perspectives to introduce an adaptive threshold for the gradient magnitude calculated according to noise level in each image and add a quantitative evaluation using a ground truth.

References

Aijazi, A. K., Checchin, P. and Trassoudaine, L. (2014), 'Automatic detection and feature estimation of windows for refining building facades in 3d urban point clouds', *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **2**(3), 1.

Akinlar, C. and Topal, C. (2011), 'EDLines: A real-time line segment detector with a false detection control', *Pattern Recognition Letters* **32**(13), 1633–1642.

Al-Durgham, K., Habib, A. and Kwak, E. (2013), 'Ransac approach for automated registration of terrestrial laser scans using linear features', *ISPRS Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci* **2**, 13–18.

Alshawa, M. (2007), 'lCL: Iterative closest line a novel point cloud registration algorithm based on linear features', *Ekscentar* (10), 53–59.

Amberg, B., Romdhani, S. and Vetter, T. (2007), Optimal step nonrigid ICP algorithms for surface registration, *in* '2007 IEEE Conference on Computer Vision and Pattern Recognition', IEEE, pp. 1–8.

Assi, R., Landes, T., Macher, H. and Grussenmeyer, P. (2019), 'Energy function algorithm for detection of openings in indoor point clouds', *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* .

Assi, R., Landes, T., Murtiyoso, A. and Grussenmeyer, P. (2019), 'Assessment of a keypoints detector for the registration of indoor and outdoor heritage point clouds', *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* .

Baghani, A., Valadan Zoej, M. J. and Mokhtarzade, M. (2018), 'Automatic hierarchical registration of aerial and terrestrial image-based point clouds', *European Journal of Remote Sensing* **51**(1), 436–456.

Baillard, C., Schmid, C., Zisserman, A. and Fitzgibbon, A. (1999), Automatic line matching and 3d reconstruction of buildings from multiple views, *in* 'ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery', Vol. 32, pp. 69–80.

Bazazian, D., Casas, J. R. and Ruiz-Hidalgo, J. (2015), Fast and robust edge extraction in unorganized point clouds, *in* '2015 international conference on digital image computing: techniques and applications (DICTA)', IEEE, pp. 1–8.

Besl, P. et McKay, N. (1992), 'A method for registration two 3-d shape bacterial availability', *IEEE Trans Pattern Analysis and Machine Intelligence* **14**, 232–256.

Borrmann, A., Koch, C. and Beetz, J. (2018), 'Building information modeling: Why? what? how?: Technology foundations and industry practice chapter 1 building information modeling–why? what? how?(february 2019)'.

Borrmann, D., Elseberg, J., Lingemann, K. and Nüchter, A. (2011), 'The 3d Hough transform for plane detection in point clouds: A review and a new accumulator design', *3D Research* **2**(2), 1–13.

Bughin, E. and Almansa, A. (2010), Planar patch detection for disparity maps, *in* 'Proc. 3DPVT', Citeseer.

Canny, J. (1986), 'A computational approach to edge detection', *IEEE Transactions on pattern analysis and machine intelligence* (6), 679–698.

Carter, J., Schmid, K., Waters, K., Betzhold, L., Hadley, B., Mataosky, R. and Halleran, J. (2012), 'An introduction to LiDAR technology, data, and applications', *NOAA Coastal Services Center* **2**.

Castagno, J. and Atkins, E. (2020), 'Polylidar3d-fast polygon extraction from 3d data', *Sensors* **20**(17), 4819.

Cheng, L., Chen, S., Liu, X., Xu, H., Wu, Y., Li, M. and Chen, Y. (2018), 'Registration of laser scanning point clouds: A review', *Sensors* **18**(5), 1641.

Cheng, X., Cheng, X., Li, Q. and Ma, L. (2018), 'Automatic registration of terrestrial and airborne point clouds using building outline features', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **11**(2), 628–638.

Coccia, S. (2021), 'Lidar terrestre à longue portée : retour d'expérience dans le contexte d'instabilités de pente'. [retreived on 07/07/2022] `https://www.researchgate.net/publication/356036507_LiDAR_terrestre_a_longue_portee_retour_d'experience_dans_le_contexte_d'instabilites_de_pente`.

Cover, T. M. (1999), *Elements of information theory*, John Wiley & Sons.

Deng, F., Hu, M. and Guan, H. (2008), 'Automatic registration between LiDAR and digital images', *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **37**, 487–490.

Desolneux, A. (2016), 'When the a contrario approach becomes generative', *International Journal of Computer Vision* **116**(1), 46–65.

Ding, M., Lyngbaek, K. and Zakhor, A. (2008), Automatic registration of aerial imagery with untextured 3d LiDAR models, *in* '2008 IEEE Conference on Computer Vision and Pattern Recognition', IEEE, pp. 1–8.

Dong, Z., Liang, F., Yang, B., Xu, Y., Zang, Y., Li, J., Wang, Y., Dai, W., Fan, H., Hyyppä, J. et al. (2020), 'Registration of large-scale terrestrial laser scanner point clouds: A review and benchmark', *ISPRS Journal of Photogrammetry and Remote Sensing* **163**, 327–342.

Dong, Z., Yang, B., Liang, F., Huang, R. and Scherer, S. (2018), 'Hierarchical registration of unordered tls point clouds based ondadatata binary shape context descriptor', **144**, 61–79.

Duda, R. O. and Hart, P. E. (1972), 'Use of the Hough transformation to detect lines and curves in pictures', *Communications of the ACM* **15**(1), 11–15.

Dumbrell, A., Woodward, G. and Bohan, D. (2019), *Advances in Ecological Research*, Elsevier Science & Technology.

Edelsbrunner, H., Kirkpatrick, D. and Seidel, R. (1983), 'On the shape of a set of points in the plane', *IEEE Transactions on information theory* **29**(4), 551–559.

Elbaz, G., Avraham, T. and Fischer, A. (2017), 3d point cloud registration for localization using a deep neural network auto-encoder, *in* 'Proceedings of the IEEE conference on computer vision and pattern recognition', pp. 4631–4640.

Fischler, M. A. and Bolles, R. C. (1981), 'Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography', *Communications of the ACM* **24**(6), 381–395.

Flamant, P. H. (2019), 'Comprendre la science du LiDAR', *Photoniques* (97), 40–44.

Forstner, W. and Khoshelham, K. (2017), Efficient and accurate registration of point clouds with plane to plane correspondences, *in* 'Proceedings of the IEEE International Conference on Computer Vision Workshops', pp. 2165–2173.

Gelfand, N., Ikemoto, L., Rusinkiewicz, S. and Levoy, M. (2003), Geometrically stable sampling for the ICP algorithm, *in* 'Fourth International Conference on 3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings.', IEEE, pp. 260–267.

Gojcic, Z., Zhou, C., Wegner, J. D., Guibas, L. J. and Birdal, T. (2020), Learning multiview 3d point cloud registration, *in* 'Proceedings of the IEEE/CVF conference on computer vision and pattern recognition', pp. 1759–1769.

Guinard, S. A., Mallé, Z., Ennafii, O., Monasse, P. and Vallet, B. (2020), 'Planar polygons detection in Lidar scans based on sensor topology enhanced RANSAC', *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences* **5**(2).

Guinard, S. and Vallet, B. (2018), 'Sensor-topology based simplicial complex reconstruction from mobile laser scanning', *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* **4**, 121–128.

Habib, A. F., Shin, S., Kim, C. and Al-Durgham, M. (2006), Integration of photogrammetric and LiDAR data in a multi-primitive triangulation environment, *in* 'Innovations in 3D Geo Information Systems', Springer, pp. 29–45.

Hackel, T., Wegner, J. D. and Schindler, K. (2016), Contour detection in unstructured 3d point clouds, *in* 'Proceedings of the IEEE conference on computer vision and pattern recognition', pp. 1610–1618.

Heuel, S. and Forstner, W. (2001), Matching, reconstructing and grouping 3d lines from multiple views using uncertain projective geometry, *in* 'Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001', Vol. 2, IEEE, pp. II–II.

Hofer, M., Maurer, M. and Bischof, H. (2017), 'Efficient 3d scene abstraction using line segments', *Computer Vision and Image Understanding* **157**, 167–178.

Hojjatoleslami, S. and Kittler, J. (1998), 'Region growing: a new approach', *IEEE Transactions on Image processing* **7**(7), 1079–1084.

Hossain, M. A. and Yeoh, J. K. (2018), Bim for existing buildings: potential opportunities and barriers, *in* 'IOP Conference Series: Materials Science and Engineering', Vol. 371, IOP Publishing, p. 012051.

Hu, L., Xiao, J. and Wang, Y. (2020), 'An automatic 3d registration method for rock mass point clouds based on plane detection and polygon matching', *The Visual Computer* **36**(4), 669–681.

Jain, A., Kurz, C., Thormählen, T. and Seidel, H.-P. (2010), Exploiting global connectivity constraints for reconstruction of 3d line segments from images, *in* '2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition', IEEE, pp. 1586–1593.

Kallisto (2022), 'Scanner 3d - laser - faro focus 3d x330'. [retreived on 08/07/2022] `https://www.kallisto.net/3-scanner-3d-laser-faro-focus-3d-x330.html`.

Kim, C., Ghanma, M. and Habib, A. (2006), 'Integration of photogrammetric and LiDAR data for realistic 3d model generation', *Department of Geomatics Engineering, University of Calgary, Canada* .

Kim, Y., Kang, B.-N. and Kim, D. (2017), Detector with focus: Normalizing gradient in image pyramid, *in* '2017 IEEE International Conference on Image Processing (ICIP)', IEEE, pp. 420–424.

Koch, T., Korner, M. and Fraundorfer, F. (2016), Automatic alignment of indoor and outdoor building models using 3d line segments, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops', pp. 10–18.

Kukko, A., Jaakkola, A. and Hyyppä, J. (2016), 'Airborne laser scanning using uass', *GIM International* **30**, 10–13.

Kukko, A., Kaartinen, H., Hyyppä, J. and Chen, Y. (2012), 'Multiplatform mobile laser scanning: Usability and performance', *Sensors* **12**(9), 11712–11733.

Kukko, A. et al. (2013), *Mobile Laser Scanning–System development, performance and applications*, Finnish Geodetic Institute.

Kumar Mishra, R. and Zhang, Y. (2012), 'A review of optical imagery and airborne LiDAR data registration methods', *The Open Remote Sensing Journal* **5**(1).

Kurczyński, Z. (2019), 'Airborne laser scanning in poland-between science and practice', *Archiwum Fotogrametrii, Kartografii i Teledetekcji* **31**.

Li, J., Xiong, B., Biljecki, F. and Schrotter, G. (2018), 'A sliding window method for detecting corners of openings from terrestrial LiDAR data', *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **42**(4/W10), 97–103.

Li, K., Yao, J., Li, L. and Liu, Y. (2016), 3d line segment reconstruction in structured scenes via coplanar line segment clustering, *in* 'Asian Conference on Computer Vision', Springer, pp. 46–61.

Limberger, F. A. and Oliveira, M. M. (2015), 'Real-time detection of planar regions in unorganized point clouds', *Pattern Recognition* **48**(6), 2043–2053.

Lin, Y., Wang, C., Chen, B., Zai, D. and Li, J. (2017), 'Facet segmentation-based line segment extraction for large-scale point clouds', *IEEE Transactions on Geoscience and Remote Sensing* **55**(9), 4839–4854.

Lin, Y., Wang, C., Cheng, J., Chen, B., Jia, F., Chen, Z. and Li, J. (2015), 'Line segment extraction for large scale unorganized point clouds', *ISPRS Journal of Photogrammetry and Remote Sensing* **102**, 172–183.

Liu, L. and Stamos, I. (2012), 'A systematic approach for 2d-image to 3d-range registration in urban environments', *Computer Vision and Image Understanding* **116**(1), 25–37.

Low, K.-L. (2004), 'Linear least-squares optimization for point-to-plane ICP surface registration', *Chapel Hill, University of North Carolina* **4**(10), 1–3.

Lowe, D. G. (2004), 'Distinctive image features from scale-invariant keypoints', *International journal of computer vision* **60**(2), 91–110.

Lu, X., Liu, Y. and Li, K. (2019), 'Fast 3d line segment detection from unorganized point cloud', *arXiv preprint arXiv:1901.02532* .

Magnusson, M., Lilienthal, A. and Duckett, T. (2007), 'Scan registration for autonomous mining vehicles using 3d-ndt', *Journal of Field Robotics* **24**(10), 803–827.

Mastin, A., Kepner, J. and Fisher, J. (2009), Automatic registration of LiDAR and optical images of urban scenes, *in* '2009 IEEE conference on computer vision and pattern recognition', IEEE, pp. 2639–2646.

Mehendale, N. and Neoge, S. (2020), 'Review on lidar technology', *SSRN Electronic Journal* .

Mehta, S., Patel, A. and Mehta, J. (2015), Ccd or cmos image sensor for photography, *in* '2015 International conference on communications and signal processing (ICCSP)', IEEE, pp. 0291–0294.

Miled, M., Soheilian, B., Habets, E. and Vallet, B. (2016), 'Hybrid online mobile laser scanner calibration through image alignment by mutual information.', *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences* **3**(1).

Mishra, V. K., Kumar, S. and Shukla, N. (2017), 'Image acquisition and techniques to perform image acquisition', *SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology* **9**(01), 21–24.

Mlekuž, D. (2018), 'Airborne laser scanning and landscape archaeology', *Opvscvla archaeologica* **39**(1), 85–95.

Moisan, L., Desolneux, A. and Morel, J. (2008), 'From gestalt theory to image analysis'.

Monnier, F., Vallet, B., Paparoditis, N., Papelard, J.-P. and David, N. (2013), 'Mise en cohérence de données laser mobile sur un modèle cartographique par recalage non-rigide', *Revue française de photogrammétrie et de télédétection* (202), 27–41.

Murtiyoso, A. and Grussenmeyer, P. (2018), Comparison and assessment of 3d registration and georeferencing approaches of point clouds in the case of exterior and interior heritage building recording, *in* 'ISPRS TC II Mid-Term Symposium" Towards Photogrammetry 2020"', Vol. 42, Copernicus Publications, pp. 745–751.

Nag, S. (2017), 'Image registration techniques: a survey', *arXiv preprint arXiv:1712.07540* .

Nan, L. and Wonka, P. (2017), Polyfit: Polygonal surface reconstruction from point clouds, *in* 'Proceedings of the IEEE International Conference on Computer Vision', pp. 2353–2361.

Nguatem, W., Drauschke, M. and Mayer, H. (2014), 'Localization of windows and doors in 3d point clouds of facades', *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* **2**(3), 87.

Paparoditis, N., Papelard, J.-P., Cannelle, B., Devaux, A., Soheilian, B., David, N. and Houzay, E. (2012), 'Stereopolis ii: A multi-purpose and multi-sensor 3d mobile mapping system for street visualisation and 3d metrology', *Revue française de photogrammétrie et de télédétection* **200**(1), 69–79.

Pfeifer, N. (2007), 'Overview of TLS systems, overall processing and applications, theory and application of laser scanning', *ISPRS summer school, Ljubljana, Solvenia* .

Pu, S. and Vosselman, G. (2007), 'Extracting windows from terrestrial laser scanning', *Intl Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* **36**, 12–14.

Rabbani, T., Dijkman, S., van den Heuvel, F. and Vosselman, G. (2007), 'An integrated approach for modelling and global registration of point clouds', *ISPRS journal of Photogrammetry and Remote Sensing* **61**(6), 355–370.

Recky, M. and Leberl, F. (2010), Window detection in complex facades, *in* '2010 2nd European Workshop on Visual Information Processing (EUVIP)', IEEE, pp. 220–225.

Roncat, A. (2016), 'The geometry of airborne laser scanning in a kinematical framework', *http://dx.doi.org/10.13140/RG.2.2.21529.83042* . Retrieved on Dec. 19, 2018.

Salaün, Y., Marlet, R. and Monasse, P. (2016), Multiscale line segment detector for robust and accurate sfm, *in* '2016 23rd International Conference on Pattern Recognition (ICPR)', IEEE, pp. 2000–2005.

Schnabel, R., Wahl, R. and Klein, R. (2007), Efficient ransac for point-cloud shape detection, *in* 'Computer graphics forum', Vol. 26, Wiley Online Library, pp. 214–226.

Sharp, G. C., Lee, S. W. and Wehe, D. K. (2002), 'ICP registration using invariant features', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(1), 90–102.

Tang, Y. and Feng, J. (2015), Hierarchical multiview rigid registration, *in* 'Computer Graphics Forum', Vol. 34, Wiley Online Library, pp. 77–87.

Teo, T.-A. and Huang, S.-H. (2014), 'Surface-based registration of airborne and terrestrial mobile LiDAR point clouds', *Remote Sensing* **6**(12), 12686–12707.

Theiler, P., Schindler, K. et al. (2012), 'Automatic registration of terrestrial laser scanner point clouds using natural planar surfaces', *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* **3**, 173–178.

Toldo, R. and Fusiello, A. (2008), Robust multiple structures estimation with J-linkage, *in* 'European conference on computer vision', Springer, pp. 537–547.

Torr, P. H. and Zisserman, A. (2000), 'MLESAC: A new robust estimator with application to estimating image geometry', *Computer vision and image understanding* **78**(1), 138–156.

Tuttas, S. and Stilla, U. (2011), 'Window detection in sparse point clouds using indoor points', *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* **38**(3).

Urbančič, T., Roškar, Ž., Kosmatin Fras, M. and Grigillo, D. (2019), 'New target for accurate terrestrial laser scanning and unmanned aerial vehicle point cloud registration', *Sensors* **19**(14), 3179.

Vincent, N., Christian, H., Fabrice, H. and Jean-Paul, A. (2013), Recalage d'images médicales, Technical report.

Von Gioi, R. G., Jakubowicz, J., Morel, J.-M. and Randall, G. (2012), 'LSD: a line segment detector', *Image Processing On Line* **2**, 35–55.

Von Hansen, W., Gross, H. and Thoennessen, U. (2008), 'Line-based registration of terrestrial and airborne LiDAR data', *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci* **37**, 161–166.

Wang, R., Bach, J. and Ferrie, F. P. (2011), Window detection from mobile LiDAR data, *in* '2011 IEEE Workshop on Applications of Computer Vision (WACV)', IEEE, pp. 58–65.

Wang, R. and Ferrie, F. P. (2015), 'Automatic registration method for mobile LiDAR data', *Optical Engineering* **54**(1), 013108.

Woo, J. H. (2006), Bim (building information modeling) and pedagogical challenges, *in* 'Proceedings of the 43rd ASC national annual conference', pp. 12–14.

Xiao, J., Adler, B. and Zhang, H. (2012), 3d point cloud registration based on planar surfaces, *in* '2012 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)', IEEE, pp. 40–45.

Xu, B., Jiang, W., Shan, J., Zhang, J. and Li, L. (2016), 'Investigation on the weighted ransac approaches for building roof plane segmentation from LiDAR point clouds', *Remote Sensing* **8**(1), 5.

Yadav, J. and Sharma, M. (2013), 'A review of k-mean algorithm', *Int. J. Eng. Trends Technol* **4**(7), 2972–2976.

Zhou, H., Peng, K., Zhou, D., Fan, W. and Liu, Y. (2020), 'Uncertainty analysis of 3d line reconstruction in a new minimal spatial line representation', *Applied Sciences* **10**(3), 1096.

Zhou, W. and Yan, H. (2014), 'Alpha shape and Delaunay triangulation in studies of protein-related interactions', *Briefings in bioinformatics* **15**(1), 54–64.

Zhu, B., Ye, Y., Yang, C., Zhou, L., Liu, H. and Cao, Y. (2020), 'Fast and robust registration of aerial images and LiDAR data based on structrual features and 3d phase correlation', *arXiv preprint arXiv:2004.09811* .